PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12P 19/34, C12Q 1/68 — A1

(11) International Publication Number: **WO 00/49169**

(43) International Publication Date: 24 August 2000 (24.08.00)

(21) International Application Number: PCT/US00/04251

(22) International Filing Date: 18 February 2000 (18.02.00)

(30) Priority Data:
60/120,787 19 February 1999 (19.02.99) US

(71) Applicant: DZGENES, L.L.C. [US/US]; 518 Bonhomme Woods Drive, St. Louis, MO 63132 (US).

(72) Inventor: MOSKOWITZ, David, W.; DzGenes, LLC, 518 Bonhomme Woods Drive, St. Louis, MO 63132 (US).

(74) Agents: BUTLER, James, E. et al.; Senniger, Powers, Leavitt & Roedel, One Metropolitan Square, 16th floor, St. Louis, MO 63102 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DIAGNOSTIC POLYMORPHISMS

(57) Abstract

Disclosed are single nucleotide polymorphisms (SMPs) associated with end stage renal disease. Also disclosed are methods for using SMPs to determine susceptibility to end stage renal disease; nucleotide sequences containing SNPs; kits for determining the presence of SNPs; and methods of treatment or prophylaxis based on the presence of SNPs.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DIAGNOSTIC POLYMORPHISMS

CROSS–REFERENCE TO RELATED APPLICATIONS

This application claims priority from provisional application 60/120,787, filed February 19, 1999, which is hereby incorporated by reference in its entirety for all purposes.

BACKGROUND

This invention relates to detection of individuals at risk for pathological conditions based on the presence of single nucleotide polymorphisms (SNPs).

During the course of evolution, spontaneous mutations appear in the genomes of organisms. It has been estimated that variations in genomic DNA sequences are created continuously at a rate of about 100 new single base changes per individual (Kondrashow, *J. Theor. Biol.*, 175:583-594, 1995; Crow, *Exp. Clin. Immunogenet.*, 12:121-128, 1995) These changes, in the progenitor nucleotide sequences, may confer an evolutionary advantage, in which case the frequency of the mutation will likely increase, an evolutionary disadvantage in which case the frequency of the mutation is likely to decrease, or the mutation will be neutral. In certain cases, the mutation may be lethal in which case the mutation is not passed on to the next generation and so is quickly eliminated from the population. In many cases, an equilibrium is established between the progenitor and mutant sequences so that both are present in the population. The presence of both forms of the sequence results in genetic variation or polymorphism. Over time, a significant number of mutations can accumulate within a population such that considerable polymorphism can exist between individuals within the population.

Numerous types of polymorphism are known to exist. Polymorphisms can be created when DNA sequences are either inserted or deleted from the genome, for example, by viral insertion. Another source of sequence variation can be caused by the presence of repeated sequences in the genome variously termed short tandem repeats (STR), variable number tandem repeats (VNTR), short sequence repeats (SSR) or microsatellites. These repeats can be dinucleotide, trinucleotide,

tetranucleotide or pentanucleotide repeats. Polymorphism results from variation in the number of repeated sequences found at a particular locus.

By far the most common source of variation in the genome are single nucleotide polymorphisms or SNPs. SNPs account for approximately 90% of human DNA polymorphism (Collins et al., *Genome Res.*, 8:1229-1231, 1998). SNPs are single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in a population. In addition, the least frequent allele must occur at a frequency of 1% or greater. Several definitions of SNPs exist in the literature (Brooks, *Gene*, 234:177-186, 1999). As used herein, the term "single nucleotide polymorphism" or "SNP" includes all single base variants and so includes nucleotide insertions and deletions in addition to single nucleotide substitutions (e.g. A->G). Nucleotide substitutions are of two types. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine for a pyrimidine or vice versa.

The typical frequency at which SNPs are observed is about 1 per 1000 base pairs (Li and Sadler, *Genetics*, 129:513-523, 1991; Wang et al., *Science*, 280:1077-1082, 1998; Harding et al., *Am. J. Human Genet.*, 60:772-789, 1997; Taillon-Miller et al., *Genome Res.*, 8:748-754, 1998). The frequency of SNPs varies with the type and location of the change. In base substitutions, two-thirds of the substitutions involve the C<->T (G<->A) type. This variation in frequency is thought to be related to 5-methylcytosine deamination reactions that occur frequently, particularly at CpG dinucleotides. In regard to location, SNPs occur at a much higher frequency in non-coding regions than they do in coding regions.

SNPs can be associated with disease conditions in humans or animals. The association can be direct, as in the case of genetic diseases where the alteration in the genetic code caused by the SNP directly results in the disease condition. Examples of diseases in which single nucleotide polymorphisms result in disease conditions are sickle cell anemia and cystic fibrosis. The association can also be indirect, where the SNP does not directly cause the disease but alters the physiological environment such that there is an increased likelihood that the patient will develop

the disease. SNPs can also be associated with disease conditions, but play no direct or indirect role in causing the disease. In this case, the SNP is located close to the defective gene, usually within 5 centimorgans, such that there is a strong association between the presence of the SNP and the disease state. Because of the high frequency of SNPs within the genome, there is a greater probability that a SNP will be linked to a genetic locus of interest than other types of genetic markers.

Disease associated SNPs can occur in coding and non-coding regions of the genome. When located in a coding region, the presence of the SNP can result in the production of a protein that is non-functional or has decreased function. More frequently, SNPs occur in non-coding regions. If the SNP occurs in a regulatory region, it may affect expression of the protein. For example, the presence of a SNP in a promoter region, may cause decreased expression of a protein. If the protein is involved in protecting the body against development of a pathological condition, this decreased expression can make the individual more susceptible to the condition.

Numerous methods exist for the detection of SNPs within a nucleotide sequence. A review of many of these methods can be found in Landegren et al., *Genome Res.*, 8:769-776, 1998. SNPs can be detected by restriction fragment length polymorphism (RFLP)(U.S. Patent Nos. 5,324,631; 5,645,995). RFLP analysis of the SNPs, however, is limited to cases where the SNP either creates or destroys a restriction enzyme cleavage site. SNPs can also be detected by direct sequencing of the nucleotide sequence of interest. Numerous assays based on hybridization have also been developed to detect SNPs. In addition, mismatch distinction by polymerases and ligases have also been used to detect SNPs.

There is growing recognition that SNPs can provide a powerful tool for the detection of individuals whose genetic make-up alters their susceptibility to certain diseases. There are four primary reasons why SNPs are especially suited for the identification of genotypes which predispose an individual to develop a disease condition. First, SNPs are by far the most prevalent type of polymorphism present in the genome and so are likely to be present in or near any locus of interest. Second, SNPs located in genes can be expected to directly affect protein structure or expression levels and so may serve not only as markers but as candidates for gene

therapy treatments to cure or prevent a disease. Third, SNPs show greater genetic stability than repeated sequences and so are less likely to undergo changes which would complicate diagnosis. Fourth, the increasing efficiency of methods of detection of SNPs make them especially suitable for high throughput typing systems necessary to screen large populations.

One disease for which the discovery of markers to detect increased genetic susceptibility is critically needed is end-stage renal disease. End-stage renal disease (ESRD) is defined as the condition when life becomes impossible without replacement of renal functions either by kidney dialysis or kidney transplantation. Hypertension (HTN) and non-insulin dependent diabetes (NIDDM) are the leading causes of end-stage renal disease (ESRD) nationally (United States Renal Data System, Table IV-3, p. 49, 1994). There is currently an epidemic of ESRD, due mainly to the aging of the American population. The ESRD epidemic is of special concern among African Americans where the incidence of ESRD is four- to six-fold higher than for Caucasians (Brancati et al., *J. Am. Med. Assoc.*, 268:3079-3084, 1992), but where treatment of hypertension, a causative factor in ESRD, is less effective (Walker et al., *J. Am. Med. Assoc.*, 268:3085-3091, 1992).

There are currently 200,000 patients with ESRD receiving renal replacement therapy (dialysis or renal transplantation), with an annual cost of $13 billion. These numbers will certainly increase as the population of the nation continues to age. Since 1980, when complete data became available for the first time, most new cases of ESRD have been ascribed to NIDDM or hypertension. The incidence of ESRD due to NIDDM or hypertension is still increasing, suggesting that the U.S. is in the early phase of an epidemic of ESRD. Preventing ESRD would save at least $30,000 per patient, per year in dialysis costs alone, as well as enhance the patient's quality of life and ability to work. It is clearly the ideal method of cost-containment for renal disease. Without effective prevention of ESRD, the nation will instead be forced to adopt less humane methods of cost-containment, such as denial of access (gate-keeping), or rely upon unrealistic expectations about patient reimbursement rates, etc.

An ideal approach to prevention of ESRD would be the identification of any genes that predispose an individual to ESRD early enough to be able to counteract this predisposition. Knowledge of ESRD-predisposing genes is essential for truly effective delay, or, ideally, prevention of ESRD.

SUMMARY

The present inventor has discovered novel single nucleotide polymorphisms (SNPs) associated with the development of end-stage renal disease in patients with either hypertension or noninsulin-dependent diabetes mellitus. As such, these polymorphisms provide a method for diagnosing a genetic predisposition for the development of end-stage renal disease in individuals. Information obtained from the detection of SNPs associated with the development of end-stage renal disease is of great value in the treatment and prevention of the disease.

Accordingly, one aspect of the present invention provides a method for diagnosing a genetic predisposition for end-stage renal disease in a subject, comprising obtaining a sample containing at least one polynucleotide from the subject, and analyzing the polynucleotide to detect a genetic polymorphism wherein said genetic polymorphism is associated with an altered susceptibility to developing end-stage renal disease. In one embodiment, the polymorphism is located in a gene selected from the group of genes in Table 1.

Another aspect of the present invention provides an isolated nucleic acid sequence comprising at least 10 contiguous nucleotides from SEQ ID NO: 1 or 2, or their complements, wherein the sequence contains at least one polymorphic site associated with a disease and in particular end-stage renal disease.

Yet another aspect of the invention is a kit for the detection of a polymorphism comprising, at a minimum, at least one polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO: 1 or 2, or their complements, wherein the polynucleotide contains at least one polymorphic site associated with end-stage renal disease.

Yet another aspect of the invention provides a method for treating end stage renal disease comprising, obtaining a sample of biological material containing at

least one polynucleotide from the subject; analyzing the polynucleotide to detect the presence of at least one polymorphism associated with end stage renal disease; and treating the subject in such a way as to counteract the effect of any such polymorphism detected.

Still another aspect of the invention provides a method for the prophylactic treatment of a subject with a genetic predisposition to end stage renal disease comprising, obtaining a sample of biological material containing at least one polynucleotide from the subject; analyzing the polynucleotide to detect the presence of at least one polymorphism associated with end stage renal disease; and treating the subject.

Further scope of the applicability of the present invention will become apparent from the detailed description and drawings provided below. It should be understood, however, that the following detailed description and examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from the following detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings where:

Figure 1 shows SEQ ID NO: 1, the nucleotide sequence of the TGF-b1 promoter region as contained in GenBank (accession no. J04431).

Figure. 2 shows SEQ ID NO: 2, the nucleotide sequence of the TGF-b1 receptor II promoter region as contained in GenBank (accession no. U37070)

DEFINITIONS

nt = nucelotide

bp = base pair

kb = kilobase; 1000 base pairs

ESRD = end-stage renal disease

HTN = hypertension

NIDDM = noninsulin-dependent diabetes mellitus

CRF = chronic renal failure

T-GF = tubulo-glomerular feedback

CRG = compensatory renal growth

MODY = maturity-onset diabetes of the young

RFLP = restriction fragment length polymorphism

MASDA = multiplexed allele-specific diagnostic assay

MADGE = microtiter array diagonal gel electrophoresis

OLA = oligonucleotide ligation assay

DOL = dye-labeled oligonucleotide ligation assay

SNP = single nucleotide polymorphism

PCR = polymerase chain reaction

"Polynucleotide" and "oligonucleotide" are used interchangeably and mean a linear polymer of at least 2 nucleotides joined together by phosphodiester bonds and may consist of either ribonucleotides or deoxyribonucleotides.

"Sequence" means the linear order in which monomers occur in a polymer, for example, the order of amino acids in a polypeptide or the order of nucleotides in a polynucleotide.

"Polymorphism" refers to a set of genetic variants at a particular genetic locus among individuals in a population.

"Promoter" means a regulatory sequence of DNA that is involved in the binding of RNA polymerase to initiate transcription of a gene. A "gene" is a segment of DNA involved in producing a peptide, polypeptide, or protein, including the coding region, non-coding regions preceding ("leader") and following ("trailer") the coding region, as well as intervening non-coding sequences ("introns") between individual coding segments ("exons"). A promoter is herein considered as a part of the corresponding gene. Coding refers to the representation of amino acids, start and stop signals in a three base "triplet" code. Promoters are often upstream ("5' to") the transcription initiation site of the gene.

"Gene therapy" means the introduction of a functional gene or genes from some source by any suitable method into a living cell to correct for a genetic defect.

"Wild type allele" means the most frequently encountered allele of a given nucleotide sequence of an organism.

"Genetic variant" or "variant" means a specific genetic variant which is present at a particular genetic locus in at least one individual in a population and that differs from the wild type.

As used herein the terms "patient" and "subject" are not limited to human beings, but are intended to include all vertebrate animals in addition to human beings.

As used herein the terms "genetic predisposition", "genetic susceptibility" and "susceptibility" all refer to the likelihood that an individual subject will develop a particular disease, condition or disorder. For example, a subject with an increased susceptibility or predisposition will be more likely that average to develop a disease, while a subject with a decreased predisposition will be less likely than average to develop the disease. A genetic variant is associated with an altered susceptibility or predisposition if the allele frequency of the genetic variant in a population or subpopulation with a disease, condition or disorder varies from its allele frequency in the population without the disease, condition or disorder (control population) or a control sequence (wild type) by at least 1%, preferably by at least 2%, more preferably by at least 4% and more preferably still by at least 8%.

As used herein "isolated nucleic acid" means a species of the invention that is the predominate species present (i.e., on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

As used herein, "allele frequency" means the frequency that a given allele appears in a population.

DETAILED DESCRIPTION

All publications, patents, patent applications and other references cited in this application are herein incorporated by reference in their entirety as if each individual publication, patent, patent application or other reference were specifically and individually indicated to be incorporated by reference.

Compensatory renal growth

Several factors are involved in compensatory renal growth. In CRF the renal cortex gets progressively thinner. Loss of renal parenchyma thus occurs selectively from the compartment in which glomeruli are located. This is the same compartment which enlarges in compensatory renal growth (Fine, *Kidney Int.*, 29:619-634, 1986. and refs. therein). As in myocardial hypertrophy, which may lead to dilated cardiomyopathy, renal hypertrophy can be the forerunner of tissue loss.

Hepatocyte nuclear factor-1a (HNF-1a) plays a role in compensatory renal growth. This transcription factor, which is induced by HNF-4a, is present in liver and other tissues, including the kidney, where it behaves as an immediate early gene in compensatory renal growth (Moskowitz and Liu *J Urology*, 154:1560-1565, 1995). HNF-1a may be an anti-apoptotic factor in the liver as well as in other tissues (pancreas, kidney, peripheral neurons, retina), akin to NF-kB, a ubiquitous anti-apoptotic transcription factor (Beg and Baltimore *Science,* 274:782-784, 1996), since "knock-out" mice for either gene fail to develop as embryos (Duncan et al. *Science,* 281:692-695, 1998). The NF-kB2 (p65, or RelA) -/- "knockout" mouse dies on embryonic day 9 after massive hepatocyte apoptosis (Beg et al., *Nature*, 376:167-170, 1995).

The regulation of intrarenal blood flow is also a factor in compensatory renal growth. This regulation is quite complex, perhaps more so than for any other organ given its extreme blood flow/mass ratio and the exquisite matching of chemical reabsorptive work with blood flow required of the kidney (Navar et al., *Physiol. Rev.*, 76:425-536, 1996). In rodent compensatory renal growth, the renal cortex (where biochemical evidence of growth is apparent within 5 min after contralateral nephrectomy [Toback et al., *J. Clin. Invest.*, 54:91-97, 1974]) undergoes vasoconstriction within 3 min after contralateral nephrectomy (Moskowitz et al.,

Renal Failure, 18:833-846, 1996). This result was unexpected, since blood flow had earlier been found to increase substantially in the renal artery within minutes after contralateral nephrectomy (Krohn et al., *J. Urol.*, 103:564-568, 1970). For blood flow to increase in the main renal artery, yet for the renal cortex to undergo vasoconstriction, suggests that renal arterial blood flow was re-directed from cortex to medulla after uninephrectomy.

This simultaneous vasoconstriction of renal cortex and vasodilatation of renal medulla implicates primarily adenosine, since there is a selective distribution of vasoconstrictive adenosine A1 receptors in renal cortical vessels, and vasodilatory adenosine A2 receptors in renal medullary vessels (Anderson RJ. Adenosine: mechanisms of renal actions, In: *Hormones, Autacoids, and the Kidney*, eds. Goldfarb S, Ziyadeh FN. Churchill Livingstone, 1991:281-296).

Adenosine is released from cells in proportion to the rate of work demanded of them and the degree of ischemia to which they are exposed. Adenosine is derived from the breakdown of ATP, which if not replenished from ADP or AMP, is further dephosphorylated to adenosine. The medullary thick ascending limb carries out the most active (ATP-requiring) ion transport in the nephron, yet has the lowest oxygen tension. The medullary thick ascending limb is therefore likely to be the major site for adenosine generation along the nephron.

Adenosine has also long been implicated as a possible mediator of tubulo-glomerular feedback (Anderson RJ. Adenosine: mechanisms of renal actions, In: *Hormones, Autacoids, and the Kidney*, eds. Goldfarb S, Ziyadeh FN. Churchill Livingstone, 1991:281-296). Tubulo-glomerular feedback is characterized by the reduction of blood flow to individual nephrons by afferent arteriolar constriction when blood flow exceeds the capacity of the nephron mass to perform the chemical work of solute reabsorption. The renal cortical vasoconstriction that occurs within 3 minutes after contralateral nephrectomy could be most easily explained by activation of tubulo-glomerular feedback.

Model for ESRD

The progression of CRF appears to involve ongoing tubulo-glomerular feedback. Bricker's "intact nephron hypothesis" (Bricker et al. *Am. J. Med.* 28:77-97, 1960) focused on the positive feedback loop which exists for remnant nephrons, unlike nephrons in healthy kidneys. Brenner expanded this concept by calling attention to the danger of elevated glomerular capillary pressure (P_{GC}) and, more generally, hyperfiltration by single nephrons. More recently, Brenner has shown that there is a threshold number of nephrons (about one million) below which progressive renal failure occurs (Brenner et al., *Kidney Intl.*, Suppl., 63:S124-S127, 1997). These observations are all compatible with the normal functioning of tubulo-glomerular feedback: called upon briefly, as after contralateral nephrectomy, compensatory renal growth (CRG) results; activated continuously, which occurs when the functioning renal mass is less than one million nephrons, tubulo-glomerular feedback leads to apoptosis. This, in turn, leads to CRF and ESRD. The body is thus apparently designed to obliterate the kidneys altogether (through apoptosis) rather than to allow glomerular filtration to exceed tubular capacity for reabsorption, which would result in rapid depletion of total body water.

This model can be visualized as follows:

Initial loss of nephrons---> T-GF ---> >10^6 nephrons remain?--->

--->YES----> CRG (STOP)

T

T

T

NO--->continuing loss of nephrons--->apoptosis---> CRF---> ESRD

Thus, factors such as adenosine and HNF-1a, which affect tubulo-glomerular feedback and (indirectly) CRG, also influence the progression to ESRD.

Genetic Predisposition to ESRD

No more than 30% of patients with NIDDM, and 10% of patients with essential hypertension, develop nephropathy, suggesting a genetic component. A genetic influence in susceptibility to renal disease is also suggested by the findings of Yamagata et al., *Nature*, 384:458-460, 1996; Krohn et al., *J. Urol.* 103:564-568, 1970; and Anderson RJ. Adenosine: mechanisms of renal actions, In: *Hormones, Autacoids, and the Kidney*, eds. Goldfarb S, Ziyadeh FN. Churchill Livingstone, 1991:281-29621. Alterations in the coding sequence of a single major gene are responsible for both NIDDM and its complications (including nephropathy) in "maturity-onset diabetes of the young" (MODY), which is inherited as an autosomal dominant disease. Nonsense mutations in the coding sequence of either of two transcription factors, HNF-1a (MODY 3) or HNF-4a (MODY 1), are sufficient to cause disease in heterozygotes.

MODY resembles other Mendelian (single-gene) disorders, such as cystic fibrosis and Huntington's disease, in its early onset (usually before the 4th decade) and its clear pattern of inheritance. Mendelian disorders, which are relatively rare in populations, have so far all involved alterations in the coding sequence of non-redundant genes.

By contrast, the median age for the incidence of ESRD due to hypertension is 68, and due to NIDDM, 61 (United States Renal Data System. Table IV-3, p. 49, 1994). Although there is clearly an increased familial risk in ESRD (Bergman et al., *Am. J. Kid. Dis.*, 27:341-346, 1996), there is no evidence for the segregation of a single major gene, as in MODY.

In the simplest case (MODY 1 and 3), NIDDM and its complications appear to result from accelerated apoptosis in the pancreas and affected target organs, including the kidney, brought on by having only ~50% of the normal activity of an anti-apoptotic factor (HNF-1a or HNF-4a) (Yamagata K et al., *Nature*, 384:455-458, 1996; Yamagata K et al., *Nature*, 384:458-460, 1996).

Based on the above observations, it is clear that hypertension and NIDDM are usually not sufficient to cause renal failure, since ESRD is a relatively rare complication of each. Also, progressive renal failure due to hypertension or

NIDDM is characterized by loss of parenchyma without inflammation (e.g. localized pain, fever, sweats). These symptoms define apoptosis (Wyllie AH. Cell death: a new classification separating apoptosis from necrosis. In: *Cell death in biology and pathology,* Wyllie AH. Ed, Chapman & Hall, 1981:9-34), further supporting a role for renal apoptosis in ESRD.

Thus, vasoconstriction occurs early in renal growth (Moskowitz et al., *Renal Failure* 18:833-846, 1996), and CRG involves the same "immediate early" genes whose excess activity triggers apoptosis in other systems, e.g. p53, RB, c-*myc,* and cyclin D1 (Moskowitz and Liu *J Urology* 154:1560-1565, 1995 and refs. therein).

The following points are consistent with the above data, support the model for ESRD disclosed above, and provide a rationale for asserting a genetic component for susceptibility to ESRD:

- Renal apoptosis, and clinical ESRD, result from overactivity of gene products along the growth pathway.
- Common, polygenic diseases, such as ESRD, involve subtle variations in the level of normal, interacting gene products, rather than dramatic alterations in the coding sequence, as occurs in single-gene disorders.
- Renal apoptosis is triggered by the same factors which trigger compensatory renal growth (CRG).

The use of candidate genes in case-control association studies is disclosed herein to identify genetic variants at polymorphic genetic loci which serve as indicators of predisposition to ESRD. The case-control association study is a preferred method to find disease susceptibility genes, due to its low cost in comparison to pedigree-based positional cloning methods. Association studies are suited for common, polygenic diseases, in which multiple interacting polymorphisms exist, none of which is either sufficient or necessary to cause disease (Risch and Merikangas, *Science*, 273:1516-1517, 1996).

The level of expression of a gene product is determined to a significant extent by its rate of transcription, which in large part depends on regulatory sequences within the promoter region. The promoter is usually located immediately upstream (5') of the transcription initiation site of the gene. Small changes in the

promoter can affect the level of expression of a gene (Spek et al. *J. Biol. Chem.*, 273:10168-10173, 1998). In addition, the ACE I/D polymorphism illustrates that the presence of an *Alu* sequence within an intron may retard transcription (see Moskowitz, Hypertension, thermotolerance, and the "African gene": an hypothesis. Clinical Experimental Hypertension: Part A. Theory and Practice 18:1-19, 1996 for refs.). *Alu* insertions are extremely common, comprising 5% of the entire human genome, and occurring on average every 10 kb.

Thus, a single nucleotide change in a regulatory element (often only 10 bp long) in the promoter, or the presence/absence of an *Alu* sequence within an intron (as in the ACE I/D polymorphism) or within the promoter region, could alter the level of expression of a protein important for growth and/or apoptosis. Such changes are expected to be more common than coding sequence variations, and provide a better substrate for the action of natural selection, because the full activity of the gene product is retained. Only the gene's regulation has changed, which may confer a sufficient selective advantage to result in a relatively high frequency of the allele in a population.

Genes Affecting ESRD-Related Pathways

Various genes are involved, directly or indirectly, in the development of ESRD. The mechanism for their effect on ESRD may be through their involvement in the development of hypertension, CRG, NIDDM, CRF, MODY, tubulo-glomerular feedback, or other conditions. A genetic variant of any of these genes may influence the expression of the genes to increase or decrease susceptibility to ESRD. While not being bound to any particular theory, the following explanations of the potential effects of various genes on the above conditions are believed to establish a basis for asserting their possible involvement in conferring susceptibility to ESRD.

Additional renal vasoconstrictors involved in triggering compensatory renal growth are angiotensin II and norepinephrine. Both have been shown to induce a similar subset of the immediate early genes which are also expressed in compensatory renal growth (Moskowitz and Liu, *J. Urology,* 154:1560-1565, 1995

and refs. therein). These may lead to ESRD rather than merely to hypertension and insulin resistance due to activation of molecules farther downstream in the pathway of renal growth and apoptosis, as described below.

Angiotensin II specifically induces TGFb-1 (Lee et al., *J. Mol. Cell. Cardiol.*, 27:2347-2357, 1995), which induces apoptosis through a nitric oxide-independent mechanism, perhaps involving NF-kB (Martin-Sanz et al., *Hepatology* 23:1200-1207, 1996). Activation of protein kinase C (PKC) by angiotensin II, and protein kinase A (PKA) by norepinephrine, stimulate NF-kB transcription (see below). Activation of both PKC and PKA occurs early during compensatory renal growth (Moskowitz and Liu, *J. Urology,* 154:1560-1565, 1995).

In addition to induction of TGFb-1, angiotensin II appears to be especially important in causing renal apoptosis through activation of endothelin-1 (ET-1). ET-1 is a more potent vasoconstrictor than angiotensin II, and has already been implicated in nephrosclerosis (Rabelink et al., *Kidney Intl.*, 50:1827-1833, 1996). Individuals who overexpress ET-1 in response to angiotensin II may develop unusually intense and long-lasting local vasoconstriction, hypoxia, and depletion of cellular ATP, resulting in apoptosis (Richter et al., *FEBS Letters*, 378:107-110, 1996).

The endothelin A receptor (ETA), which mediates vasoconstriction, could affect ESRD because of its possible contribution to ET-1 signaling overactivity (Benigni et al., *Mineral Electrolyte Metab.*, 21: 283-291, 1995). Genetic variations in the promoter region, introns (for presence/absence of *Alu* sequence), and coding sequence of ETA can thus confer increased basal or stimulated activity to the receptor.

The transforming growth factor-b system is also a likely mediator of renal apoptosis. TGF-b is intimately connected with glomerular sclerosis, mesangial matrix expansion, and tubulointerstitial fibrosis in experimental rodent models and human glomerulnephritis (Border et al., *Kidney Intl.,* 47(Suppl 49):S-59-S-61, 1995). Of the three isoforms available, TGF-b1 has been implicated most consistently in pathologic fibrosis (Khalil et al., *Am. J. Respir. Cell. Mol. Biol.,* 14:131-138, 1996). Activation of protein kinase C early during CRG would have

the effect of stimulating TGF-b1 production, since the TGF-b1 promoter contains AP-1 sites (Kim et al., *J. Biol. Chem.*, 264:402-408, 1989). Angiotensin II has been shown to induce TGF-b1 expression in renal mesangial cells, endothelial cells, and proximal tubular epithelial cells. Thus, greater induction of TGF-b1, or greater expression of its two main receptors (TGFb-RI and TGFb-RII), may occur in patients who progress to ESRD compared to patients who never develop CRF.

Since the coding sequence of TGF-b1 is identical between mouse and human, a period of evolutionary divergence of greater than 100 hundred million years, no human polymorphisms in the coding sequence are expected. Thus the TGF-b1 promoter and introns would be more likely candidates for genetic variants than the exons of the TGF-b1 structural gene. The promoter sequences and the structural genes for TGFb-RI and TGFb-RII are also likely candidates for genetic variations.

The L-type Ca^{2+} channel may be involved in hypertension, especially among African Americans, since ACE inhibitors are not as effective in blood pressure control for African Americans as Ca^{2+} channel blockers directed against the L-type calcium channel. The a subunit, a large gene product consisting of 46 exons, requires the much smaller b subunit for activity. The b subunit (encoded by the CACNLB3 gene) has a regulatory role, and undergoes serine/threonine phosphorylation to activate the calcium channel. By analogy with other transport proteins, the b subunit likely acts as a gate to occlude the ion channel located in the a subunit (Lacinova, et al., *FEBS Letters*, 373:103-107, 1995).

Phosphorylation of the b subunit probably opens the gate by creating electrostatic charge repulsion between a negatively charged phosphate (on a serine or threonine hydroxyl group in the b subunit) and a negatively charged amino acid (e.g. aspartate or glutamate) in the a subunit which is normally hydrogen-bonded to the serine/threonine -OH group. A change in the coding sequence of the phosphorylated serine or threonine residue to a negatively charged amino acid (e.g. asp or glu, which requires a change of at least 2 nucleotides in the CACNLB3 gene) would create a constitutively open Ca^{2+} channel. Other amino acid changes in the coding sequence that destabilize the hydrogen bonding of the ser/thr hydroxyl group

might result in a less dramatic, although still supranormal, activation of the L-type Ca^{2+} channel in response to agonists. Thus, genetic variants in the CACNLB3 gene could lead to hypertension and could therefore influence susceptibility to ESRD.

The consequences of excessive Ca^{2+} entry are expected to include increased systemic vascular resistance from vascular smooth muscle cell contraction, insulin resistance, as well as enhanced apoptosis, since cell death is associated with accumulation of intracellular, especially intramitochondrial, calcium (Schwartzman and Cidlowski *Endocrine Rev.*, 14:133-151, 1993). Indeed, a major inhibitor of apoptosis, the bcl-2 gene product, binds calcium at the nuclear and mitochondrial membranes, and is active in the kidney.

Impaired bcl-2 function has been shown to be important in mouse and human models of renal cystic disease. Bcl-2 -/- "knockout" mice unexpectedly demonstrated polycystic kidney disease, associated with accelerated apoptosis during nephrogenesis (Sorenson et al., *Am. J. Physiol.* 271:F184-F193, 1996). Apoptosis is increased in human autosomal dominant polycystic kidney disease, as well as a murine model of polycystic kidney disease (Trudel et al., *J. Am. Soc. Nephrol.* 7:1624, 1996, abst.); c-*myc* is overexpressed in polycystic kidney disease in both species. Variations in the 5' promoter region and introns are thus consistent with reduced expression of bcl-2, and therefore enhanced apoptosis, in patients with ESRD.

The genes bcl-1 (encoding cyclin D1), c-*myc*, p53, Rb, WT-1, and DR1 have been found to behave as "immediate early" genes in CRG (Moskowitz and Liu, *J. Urology*, 154:1560-1565, 1995). Additionally, vHL, the von Hippel-Lindau tumor suppressor, is analogous to WT-1, the Wilm's tumor suppressor. Where known, the promoters of these tumor suppressors contain AP-1 sites, which are sensitive to protein kinase C activation (Moskowitz and Liu, *J. Urology*, 154:1560-1565, 1995). Cyclin D1 and c-*myc* are positive regulators of cellular proliferation, whereas p53, Rb, WT-1, DR1, and vHL are negative growth control elements. Where tested, these gene products have been shown to induce apoptosis when overexpressed (Steller, *Science*, 267:1445-1449, 1995). Genetic variants of promoters or introns for each of these genes may lead to overexpression of the gene product. As with

previously discussed genes, these genetic variants could take the form of single nucleotide additions, deletions, or substitutions, or the insertion or deletion of *Alu* elements. For example, p53 is known to contain an *Alu* sequence in intron 6 (GenBank Accession no. U38671), which might be deleted in ESRD patients, leading to its overexpression.

Tumor necrosis factor-a (TNFa) is expressed by phagocytes, including renal mesangial cells. When bound to its receptor (TNF receptor), apoptotic death of the TNFa-bearing cell results. Fas ligand (FasL) is similar to TNFa, although when FasL binds to its receptor (Fas), apoptotic death of the Fas-bearing cell results. Both Fas and FasL are expressed in human kidney (Gonzalez-Cuadrado et al., *Kidney Intl.*, 49:1064-1070, 1996), and TNFa has been shown to be involved in human glomerulonephritis. Here, we hypothesize that these death signals mediate apoptosis in progressive CRF due to hypertension or NIDDM. Support for an immunological role in CRF comes from the finding of specific HLA associations with hypertensive ESRD (Freedman et al., *Am. J Hypertens.*, 4:393-398, 1991). The promoter, exons, and intron/exon boundaries are known for TNFa and Fas, although not for FasL (see Table 1). Furthermore, a genetic variant has already been described in TNFa's intron 1 (Wilson et al., *Human Molec. Genetics* 1:353, 1992; GenBank accession no. L11698).

Bax, a major antagonist of bcl-2, promotes apoptosis in lymphocytes (Miyashita and Reed, *Cell*, 80:293-299, 1995). Promoter and intronic variation (e.g. single nucleotide changes or deletions of an *Alu* sequence) might increase the transcription/translation of this gene product. Bak and Bak-2 similarly interact with bcl-2 to promote apoptosis. Although none of these three gene products has yet been shown to function in the kidney, loss of bcl-2 activity leads to polycystic kidney disease (see above).

Nitric oxide (NO) has been strongly implicated in apoptosis of endothelial (Bonfoco et al., *Proc. Natl. Acad. Sci. USA*, 92:7162-7166, 1995) and vascular smooth muscle cells (Nishio et al., *Biochem. Biophys. Res. Commun.*, 221:163-168, 1996). Nitric oxide, which is vasodilatory, antagonizes the vasoconstrictive effects of angiotensin II and endothelins. Since angiotensin II promotes renal injury, nitric

oxide may protect against renal injury from systemic disease such as hypertension and non-insulin dependent diabetes mellitus (NIDDM; Bataineh and Raij, *Kidney Int.*, Suppl., 68:S140S19, 1998) Nitric oxide has also been implicated in the progression of renal disease in rats (Brooks and Contino, *Pharmacology*, 56:257-261, 1998) and humans (Noris and Remuzzi, *Contrib. Nephrol.* 119:8-15, 1996; Kone, *Am. J. Kidney Dis.*, 30:311-333, 1997; Aiello et al., *Kidney Int.*, Suppl., 65:S63-S67, 1998; Raij, *Hypertension*, 31:189-193, 1998). The nitric oxide synthase genes are recognized candidate genes for hypertension, renal failure, and cardiovascular in general (Soubrier, *Hypertension*, 31:189-193, 1998)

NO can directly oxidize (and activate) thiol-containing proteins such as NF-kB and AP-1 (Stamler, *Cell*, 78:931-936, 1994). NO can either promote apoptosis or prevent it. Above a threshold concentration, NO seems to stimulate apoptosis (Bonfoco et al., *Proc. Natl. Acad. Sci. USA*, 92:7162-7166, 1995; Stamler, *Cell*, 78:931-936, 1994).

The highest amount of NO is made by the inducible NO synthase (iNOS, NOS II), which is fully active at the prevailing intracellular calcium concentration (Ca_i ~100 nM), and, once induced, remains active for days, producing nanomolar amounts of NO (Yu et al., *Proc. Natl. Acad. Sci. USA*, 91:1691-1695, 1994). The *cis* regulatory sequences for iNOS are not fully known. However, a region of 1798 nucleotides (nt) immediately upstream (5') of the gene has been sequenced. Additional regulatory regions far upstream have been found in the human iNOS gene (de Vera ME et al., *Proc. Natl. Acad. Sci. USA*, 93:1054-1059, 1996), but have not yet been reported. Increased inducibility of iNOS would have conferred an important selection advantage, since iNOS is thought to be the major mechanism for immune cell-mediated killing of infectious agents such as parasites (e.g. malaria), bacteria, and viruses.

An additional source of renal NO is endothelial constitutive NOS (ecNOS, NOS III). ecNOS requires an elevation of Ca_i to be active, since it must bind calmodulin for activity. ecNOS, which produces picomolar amounts of NO, may thus seem an unlikely source of large amounts of NO, but it is specifically activated by shear stress (Awolesi et al., *Surgery*, 116:439-445, 1994), and may be involved in

arterial remodeling. Like endothelin-1, ecNOS may therefore account for the clinical observation that the rate of progression of CRF is proportional to the degree of hypertension. Single nucleotide variations in the 5' promoter region (1600 nt) of ecNOS might thus allow for increased induction.

Finally, NF-kB and its repressor, IkB, play an important role in apoptosis of immune cells, including monocyte/macrophage/mesangial cells, and tumor cells. NF-kB is a ubiquitous transcription factor which is present in renal tissue (Amoah-Apraku et al., *Kidney Intl.*, 48:674-682, 1995). NF-kB binding sites exist in the promoter region of many genes, including pro-apoptotic molecules such as TNFa, Fas, iNOS, TGFb-1, and anti-apoptotic gene products such as bcl-2. In this regard it is interesting that the production of TNF-b parallels the progression of CRF (Descamps-Latshca et al., *J Immunol.*, 154:882-892, 1995), and is perhaps responsible for the cachexia of ESRD. NF-kB prevents apoptosis in tumor cell lines, but may induce apoptosis in immune cells. Its net effect may depend on the relative amounts of pro- and anti-apoptotic molecules which it induces in each tissue.

NF-kB is a heterodimer, composed of p50 (NF-kB1) which is not required for activity (Baeuerle, *Biochim. Biophys. Acta.* 1072:63-80, 1991), and p65 (also called RelA, or NF-kB2), which binds to kB sites. p65 (NF-kB2) is inducible by AP-1. It is therefore not surprising that NF-kB2 has been shown to be induced by angiotensin II, an activator of PKC.

Activation of NF-kB also occurs post-translationally by phosphorylation of its repressor, IkB, on serine residues by a serine/threonine protein kinase such as PKC, PKA, or Ca^{2+}/calmodulin-dependent protein kinase. A nuclear localization signal is then exposed on NF-kB, allowing it to translocate to the nucleus, where it binds to kB sites in DNA. Altering the molar ratio between NF-kB and its cytoplasmic inhibitor, IkB, could thus lead to a state of "constitutive" activation (too little IkB) or repression (excess IkB) of NF-kB. Interestingly, IkB contains kB binding sites within its promoter sequence, establishing an efficient negative feedback loop. Loss of kB binding sites within the IkB promoter would thus lead to runaway NF-kB activity.

A summary of the human genomic sequences which may be examined is presented below. This list is not meant to be exclusive; determination of genetic variants of any other sequences which may have similar influences on ESRD, hypertension, MODY, NIDDM, CRG, CRF, or apoptosis is contemplated as within the scope of this invention.

Table 1. Summary of Genes Affecting ESRD

	Promoter (nt)	No.of Exons	Accession Nos.
Vasoconstrictors			
Endothelin-1	3576	5	J05008
Endothelin receptor-A	476	8(4114nt)*	D11144
Probable triggers for renal apoptosis			
TGFb1	1140	7	X05839, Y00112
TGFB-RI	1397	(not known)	U51139
TGFB-RII	1883	7(1765nt)*	U37070
Cyclin D1(bcl-1)	3015	(not known)	Z29078
DR1	323	(not known)	T29750
c-myc	2462	3	D10493, D90467
p53	532	11	J04238, M22881-98
RB	900	27	X16439, M27845-66
WT-1	1430	(not known)	X74840
vHL	715	(not known)	U19763
CACNLB3	(not known)	13 (1793nt)*	D43701-4
bcl-2	1394	(not known)	X51898
Implicated in renal apoptosis			
TNFa	1178	4	X021910, L11698
Fas	2344	9	X87625, X82279-86
IkBa	506	(not known)	S67380
NF-kB2 (p65, RelA)	1396	10	L01459, Z22949-54
Involved in lymphocyte apoptosis; possibly involved in renal apoptosis			
Bak	1949	3	U23765, D88396-7
Bax	972	(not known)	U17193
Bak-2	3279	1	U16812
iNOS	1798	11	U15666, U65689-99
ecNOS	600	(not known)	AF032908 (see also Marsden et al. (1993) J. Biol. Chem. 268:17478-1748

Novel Polymorphisms

The present application provides ten single nucleotide polymorphisms (SNPs) in genes associated with end stage renal diesase. Two substitution SNPs are located in the TGF-b1 promoter. An addition six SNPs, three substitutions, two

deletions, and one addition, have been found in the TGF-B-RII promoter. The location of these SNPs as well as the wild type and variant nucleotides are summarized in Table 2.

Preparation of Samples

The presence of genetic variants in the above genes or their control regions, or in any other genes that may affect susceptibility to ESRD is determined by screening nucleic acid sequences from a population of individuals for such variants. The population is preferably comprised of some individuals with ESRD, so that any genetic variants that are found can be correlated with ESRD. The population is also preferably comprised of some individuals that have known risk for ESRD, such as individuals with hypertension, NIDDM, or CRF. The population should preferably be large enough to have a reasonable chance of finding individuals with the sought-after genetic variant. As the size of the population increases, the ability to find significant correlations between a particular genetic variant and susceptibility to ESRD also increases. Preferably, the population should have 10 or more individuals.

The nucleic acid sequence can be DNA or RNA. For the assay of genomic DNA, virtually any biological sample containing genomic DNA (e.g. not pure red blood cells) can be used. For example, and without limitation, genomic DNA can be conveniently obtained from whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal cells, skin or hair. For assays using cDNA or mRNA, the target nucleic acid must be obtained from cells or tissues that express the target sequence. One preferred source and quantity of DNA is 10 to 30 ml of anticoagulated whole blood, since enough DNA can be extracted from leukocytes in such a sample to perform many repetitions of the analysis contemplated herein.

Many of the methods described herein require the amplification of DNA from target samples. This can be accomplished by any method known in the art but preferably is by the polymerase chain reaction (PCR). Optimization of conditions for conducting PCR must be determined for each reaction and can be accomplished without undue experimentation by one of ordinary skill in the art. In general, methods for conducting PCR can be found in U.S. Patent Nos 4,965,188, 4,800,159, 4,683,202, and 4,683,195; Ausbel et al., eds., *Short Protocols in Molecular Biology*, 3rd ed., Wiley, 1995; and Innis et al., eds., *PCR Protocols*, Academic Press, 1990.

Other amplification methods include the ligase chain reaction (LCR) (see, Wu and Wallace, *Genomics*, 4:560-569, 1989; Landegren et al., *Science*, 241:1077-1080, 1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA*, 86:1173-1177, 1989), self-sustained sequence replication (Guatelli et al., *Proc. Natl. Acad. Sci. USA*, 87:1874-1878, 1990), and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produces both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

Detection of Polymorphisms

Detection of Unknown Polymorphisms

Two types of detection are contemplated within the present invention. The first type involves detection of unknown SNPs by comparing nucleotide target sequences from individuals in order to detect sites of polymorphism. If the most common sequence of the target nucleotide sequence is not known, it can be determined by analyzing individual humans, animals or plants with the greatest diversity possible. Additionally the frequency of sequences found in subpopulations characterized by such factors as geography or gender can be determined.

The presence of genetic variants and in particular SNPs is determined by screening the DNA and/or RNA of a population of individuals for such variants. If it is desired to detect variants associated with a particular disease or pathology, the population is preferably comprised of some individuals with the disease or pathology, so that any genetic variants that are found can be correlated with the disease of interest. It is also preferable that the population be composed of individuals with known risk factors for the disease. The populations should preferably be large enough to have a reasonable chance to find correlations between a particular genetic variant and susceptibility to the disease of interest. In one embodiment, the population should have at least 10 individuals, in another embodiment, the population should have 40 individuals or more. In one embodiment, the population is preferably comprised of individuals who have known risk factors for ESRD such as individuals with hypertension, NIDDM, or CRF. In addition, the allele frequency of the genetic variant in a population or subpopulation with the disease or pathology should vary from its allele frequency in the population

without the disease or pathology (control population) or the control sequence (wild type) by at least 1%, preferably by at least 2%, more preferably by at least 4% and more preferably still by at least 8%.

Determination of unknown genetic variants, and in particular SNPs, within a particular nucleotide sequence among a population may be determined by any method known in the art, for example and without limitation, direct sequencing, restriction length fragment polymorphism (RFLP), single-strand conformational analysis (SSCA), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis (HET), chemical cleavage analysis (CCM) and ribonuclease cleavage.

Methods for direct sequencing of nucleotide sequences are well known to those skilled in the art and can be found for example in Ausubel et al., eds., *Short Protocols in Molecular Biology*, 3rd ed., Wiley, 1995 and Sambrook et al., *Molecular Cloning*, 2nd ed., Chap. 13, Cold Spring Harbor Laboratory Press, 1989. Sequencing can be carried out by any suitable method, for example, dideoxy sequencing (Sanger et al., *Proc. Natl. Acad. Sci. USA,* 74:5463-5467, 1977), chemical sequencing (Maxam and Gilbert, *Proc. Natl. Acad. Sci. USA*, 74:560-564, 1977) or variations thereof. Direct sequencing has the advantage of determining variation in any base pair of a particular sequence.

RFLP analysis (see, e.g. U.S. Patents No. 5,324,631 and 5,645,995) is useful for detecting the presence of genetic variants at a locus in a population when the variants differ in the size of a probed restriction fragment within the locus, such that the difference between the variants can be visualized by electrophoresis. Such differences will occur when a variant creates or eliminates a restriction site within the probed fragment. RFLP analysis is also useful for detecting a large insertion or deletion within the probed fragment. Thus, RFLP analysis is useful for detecting, e.g., an *Alu* sequence insertion or deletion in a probed DNA segment.

Single-strand conformational polymorphisms (SSCPs) can be detected in <220 bp PCR amplicons with high sensitivity (Orita et al, *Proc. Natl. Acad. Sci. USA,* 86:2766-2770, 1989; Warren et al., In: *Current Protocols in Human Genetics*, Dracopoli et al., eds, Wiley, 1994, 7.4.1-7.4.6.). Double strands are first heat-denatured. The single strands are then subjected to polyacrylamide gel electrophoresis under non-denaturing conditions at constant temperature (i.e. low voltage and long run times) at two different temperatures, typically 4-10°C and 23°C (room temperature). At low temperatures (4-10°C), the secondary structure of short

single strands (degree of intrachain hairpin formation) is sensitive to even single nucleotide changes, and can be detected as a large change in electrophoretic mobility. The method is empirical, but highly reproducible, suggesting the existence of a very limited number of folding pathways for short DNA strands at the critical temperature. Polymorphisms appear as new banding patterns when the gel is stained.

Denaturing gradient gel electrophoresis (DGGE) can detect single base mutations based on differences in migration between homo- and heteroduplexes (Myers et al., *Nature*, 313:495-498, 1985). The DNA sample to be tested is hybridized to a labeled wild type probe. The duplexes formed are then subjected to electrophoresis through a polyacrylamide gel that contains a gradient of DNA denaturant parallel to the direction of electrophoresis. Heteroduplexes formed due to single base variations are detected on the basis of differences in migration between the heteroduplexes and the homoduplexes formed.

In heteroduplex analysis (HET) (Keen et al., *Trends Genet.* 7:5, 1991), genomic DNA is amplified by the polymerase chain reaction followed by an additional denaturing step which increases the chance of heteroduplex formation in heterozygous individuals. The PCR products are then separated on Hydrolink gels where the presence of the heteroduplex is observed as an additional band.

Chemical cleavage analysis (CCM) is based on the chemical reactivity of thymine (T) when mismatched with cytosine, guanine or thymine and the chemical reactivity of cytosine (C) when mismatched with thymine, adenine or cytosine (Cotton et al., *Proc. Natl. Acad. Sci. USA*, 85:4397-4401, 1988). Duplex DNA formed by hybridization of a wild type probe with the DNA to be examined, is treated with osmium tetroxide for T and C mismatches and hydroxylamine for C mismatches. T and C mismatched bases that have reacted with the hydroxylamine or osmium tetroxide are then cleaved with piperidine. The cleavage products are then analyzed by gel electrophoresis.

Ribonuclease cleavage involves enzymatic cleavage of RNA at a single base mismatch in an RNA:DNA hybrid (Myers et al., *Science* 230:1242-1246, 1985). A ^{32}P labeled RNA probe complementary to the wild type DNA is annealed to the test DNA and then treated with ribonuclease A. If a mismatch occurs, ribonuclease A will cleave the RNA probe and the location of the mismatch can then be determined by size analysis of the cleavage products following gel electrophoresis.

Detection of Known Polymorphisms

The second type of polymorphism detection involves determining which form of a known polymorphism is present in individuals for diagnostic or epidemiological purposes. In addition to the already discussed methods for detection of polymorphisms, several methods have been developed to detect known SNPs. Many of these assays have been reviewed by Landegren et al., *Genome Res.*, 8:769-776, 1998 and will only be briefly reviewed here.

One type of assay has been termed an array hybridization assay, an example of which is the multiplexed allele-specific diagnostic assay (MASDA) (U.S. Patent No. 5,834,181; Shuber et al., *Hum. Molec. Genet.*, 6:337-347, 1997). In MASDA, samples from multiplex PCR are immobilized on a solid support. A single hybridization is conducted with a pool of labeled allele specific oligonucleotides (ASO). Any ASOs that hybridize to the samples are removed from the pool of ASOs. The support is then washed to remove unhybridized ASOs remaining in the pool. Labeled ASOs remaining on the support are detected and eluted from the support. The eluted ASOs are then sequenced to determine the mutation present.

Two assays depend on hybridization-based allele-discrimination during PCR. The TaqMan assay (U.S. Patent No. 5,962,233; Livak et al., *Nature Genet.*, 9:341-342, 1995) uses allele specific (ASO) probes with a donor dye on one end and an acceptor dye on the other end, such that the dye pair interact via fluorescence resonance energy transfer (FRET). A target sequence is amplified by PCR modified to include the addition of the labeled ASO probe. The PCR conditions are adjusted so that a single nucleotide difference will effect binding of the probe. Due to the 5' nuclease activity of the *Taq* polymerase enzyme, a perfectly complementary probe is cleaved during the PCR while a probe with a single mismatched base is not cleaved. Cleavage of the probe dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence.

An alternative to the TaqMan assay is the molecular beacons assay (U.S. Patent No. 5,925,517; Tyagi et al., *Nature Biotech.*, 16:49-53, 1998). In the molecular beacons assay, the ASO probes contain complementary sequences flanking the target specific species so that a hairpin structure is formed. The loop of the hairpin is complimentary to the target sequence while each arm of the hairpin contains either donor or acceptor dyes. When not hybridized to a donor sequence, the hairpin structure brings the donor and acceptor dye close together thereby

extinguishing the donor fluorescence. When hybridized to the specific target sequence, however, the donor and acceptor dyes are separated with an increase in fluorescence of up to 900 fold. Molecular beacons can be used in conjunction with amplification of the target sequence by PCR and provide a method for real time detection of the presence of target sequences or can be used after amplification.

High throughput screening for SNPs that affect restriction sites can be achieved by Microtiter Array Diagonal Gel Electrophoresis (MADGE) (Day and Humphries, *Anal. Biochem.*, 222:389-395, 1994). In this assay restriction fragment digested PCR products are loaded onto stackable horizontal gels with the wells arrayed in a microtiter format. During electrophoresis, the electric field is applied at an angle relative to the columns and rows of the wells allowing products from a large number of reactions to be resolved.

Additional assays for SNPs depend on mismatch distinction by polymerases and ligases. The polymerization step in PCR places high stringency requirements on correct base pairing of the 3' end of the hybridizing primers. This has allowed the use of PCR for the rapid detection of single base changes in DNA by using specifically designed oligonucleotides in a method variously called PCR amplification of specific alleles (PASA) (Sommer et al., *Mayo Clin. Proc.*, 64:1361-1372 1989; Sarker et al., *Anal. Biochem.* 1990), allele-specific amplification (ASA), allele-specific PCR, and amplification refractory mutation system (ARMS) (Newton et al., *Nuc. Acids Res.*, 1989; Nichols et al., *Genomics*, 1989; Wu et al., *Proc. Natl. Acad. Sci. USA*, 1989). In these methods, an oligonucleotide primer is designed that perfectly matches one allele but mismatches the other allele at or near the 3' end. This results in the preferential amplification of one allele over the other. By using three primers that produce two differently sized products, it can be determined whether an individual is homozygous or heterozygous for the mutation (Dutton and Sommer, *BioTechniques*,11:700-702, 1991). In another method, termed bi-PASA, four primers are used; two outer primers that bind at different distances from the site of the SNP and two allele specific inner primers (Liu et al., *Genome Res.*, 7:389-398, 1997). Each of the inner primers have a non-complementary 5' end and form a mismatch near the 3' end if the proper allele is not present. Using this system, zygosity is determined based on the size and number of PCR products produced.

The joining by DNA ligases of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially

at the 3' end. This sensitivity has been utilized in the oligonucleotide ligation assay (Landegren et al., *Science*, 241:1077-1080, 1988) and the ligase chain reaction (LCR; Barany, *Proc. Natl. Acad. Sci. USA*, 88:189-193, 1991). In OLA, the sequence surrounding the SNP is first amplified by PCR, whereas in LCR, genomic DNA can by used as a template.

In one method for mass screening for SNPs based on the OLA, amplified DNA templates are analyzed for their ability to serve as templates for ligation reactions between labeled oligonucleotide probes (Samotiaki et al., *Genomics*, 20:238-242, 1994). In this assay, two allele-specific probes labeled with either of two lanthanide labels (europium or terbium) compete for ligation to a third biotin labeled phosphorylated oligonucleotide and the signals from the allele specific oligonucleotides are compared by time-resolved fluorescence. After ligation, the oligonucleotides are collected on an avidin-coated 96-pin capture manifold. The collected oligonucleotides are then transferred to microtiter wells in which the europium and terbium ions are released. The fluorescence from the europium ions is determined for each well, followed by measurement of the terbium fluorescence.

In alternative gel-based OLA assays, numerous SNPs can be detected simultaneously using multiplex PCR and multiplex ligation (U.S. Patent No. 5,830,711; Day et al., *Genomics*, 29:152-162, 1995; Grossman et al., *Nuc. Acids Res.*, 22:4527-4534, 1994). In these assays, allele specific oligonucleotides with different markers, for example, fluorescent dyes, are used. The ligation products are then analyzed together by electrophoresis on an automatic DNA sequencer distinguishing markers by size and alleles by fluorescence. In the assay by Grossman et al., 1994, mobility is further modified by the presence of a non-nucleotide mobility modifier on one of the oligonucleotides.

A further modification of the ligation assay has been termed the dye-labeled oligonucleotide ligation (DOL) assay (U.S. Patent No. 5,945,283; Chen et al., *Genome Res.*, 8:549-556, 1998). DOL combines PCR and the oligonucleotide ligation reaction in a two-stage thermal cycling sequence with fluorescence resonance energy transfer (FRET) detection. In the assay, labeled ligation oligonucleotides are designed to have annealing temperatures lower than those of the amplification primers. After amplification, the temperature is lowered to a temperature where the ligation oligonucleotides can anneal and be ligated together. This assay requires the use of a thermostable ligase and a thermostable DNA

polymerase without 5' nuclease activity. Because FRET occurs only when the donor and acceptor dyes are in close proximity, ligation is inferred by the change in fluorescence.

In another method for the detection of SNPs termed minisequencing, the target-dependent addition by a polymerase of a specific nucleotide immediately downstream (3') to a single primer is used to determine which allele is present (U.S Patent No. 5,846,710). Using this method, several SNPs can be analyzed in parallel by separating locus specific primers on the basis of size via electrophoresis and determining allele specific incorporation using labeled nucleotides.

Determination of individual SNPs using solid phase minisequencing has been described by Syvanen et al., *Am. J. Hum. Genet.*, 52:46-59, 1993. In this method the sequence including the polymorphic site is amplified by PCR using one amplification primer which is biotinylated on its 5' end. The biotinylated PCR products are captured in streptavidin-coated microtitration wells, the wells washed, and the captured PCR products denatured. A sequencing primer is then added whose 3' end binds immediately prior to the polymorphic site, and the primer is elongated by a DNA polymerase with one single labeled dNTP complementary to the nucleotide at the polymorphic site. After the elongation reaction, the sequencing primer is released and the presence of the labeled nucleotide detected. Alternatively, dye labeled dideoxynucleoside triphosphates (ddNTPs) can be used in the elongation reaction (U.S. Patent No. 5,888,819; Shumaker et al., *Human Mut.*, 7:346-354, 1996). In this method, incorporation of the ddNTP is determined using an automatic gel sequencer.

Minisequencing has also been adapted for use with microarrays (Shumaker et al., *Human Mut.*, 7:346-354, 1996). In this case, elongation (extension) primers are attached to a solid support such as a glass slide. Methods for construction of oligonucleotide arrays are well known to those of ordinary skill in the art and can be found, for example, in *Nature Genetics*, Suppl., Vol. 21, January, 1999. PCR products are spotted on the array and allowed to anneal. The extension (elongation) reaction is carried out using a polymerase, a labeled dNTP and noncompeting ddNTPs. Incorporation of the labeled dNTP is then detected by the appropriate means. In a variation of this method suitable for use with multiplex PCR, extension is accomplished with the use of the appropriate labeled ddNTP and unlabeled ddNTPs (Pastinen et al., *Genome Res.*, 7:606-614, 1997).

Solid phase minisequencing has also been used to detect multiple polymorphic nucleotides from different templates in an undivided sample (Pastinen et al., *Clin. Chem.*, 42:1391-1397, 1996). In this method, biotinylated PCR products are captured on the avidin-coated manifold support and rendered single stranded by alkaline treatment. The manifold is then placed serially in four reaction mixtures containing extension primers of varying lengths, a DNA polymerase and a labeled ddNTP, and the extension reaction allowed to proceed. The manifolds are inserted into the slots of a gel containing formamide which releases the extended primers from the template. The extended primers are then identified by size and fluorescence on a sequencing instrument.

Fluorescence resonance energy transfer (FRET) has been used in combination with minisequencing to detect SNPs (U.S. Patent No. 5,945,283; Chen et al., *Proc. Natl. Acad. Sci. USA*, 94:10756-10761, 1997). In this method, the extension primers are labeled with a fluorescent dye, for example fluorescein. The ddNTPs used in primer extension are labeled with an appropriate FRET dye. Incorporation of the ddNTPs is determined by changes in fluorescence intensities.

The above discussion of methods for the detection of SNPs is exemplary only and is not intended to be exhaustive. Those of ordinary skill in the art will be able to envision other methods for detection of SNPs that are within the scope and spirit of the present invention.

In one embodiment the present invention provides a method for diagnosing a genetic predisposition for a disease and in particular, end-stage renal disease and hypertension. In this method, a biological sample is obtained from a subject. The subject can be a human being or any vertebrate animal. The biological sample must contain polynucleotides and preferably genomic DNA. Samples that do not contain genomic DNA, for example, pure samples of mammalian red blood cells, are not suitable for use in the method. The form of the polynucleotide is not critically important such that the use of DNA, cDNA, RNA or mRNA is contemplated within the scope of the method. The polynucleotide is then analyzed to detect the presence of a genetic variant where such variant is associated with an increased risk of developing a disease, condition or disorder, and in particular end-stage renal disease. In one embodiment, the genetic variant is located at one of the polymorphic sites contained in Table 2. In another embodiment, the genetic variant is one of the variants contained in Table 2 or the complement of any of the variants contained in

Table 2. Any method capable of detecting a genetic variant, including any of the methods previously discussed, can be used. Suitable methods include, but are not limited to, those methods based on sequencing, mini sequencing, hybridization, restriction fragment analysis, oligonucleotide ligation, or allele specific PCR.

The present invention is also directed to an isolated nucleic acid sequence of at least 10 contiguous nucleotides from SEQ ID NO: 1 or 2, or the complements of SEQ ID NO 1 or 2. In one preferred embodiment, the sequence contains at least one polymorphic site associated with a disease, and in particular end-stage renal disease. In one embodiment, the polymorphic site is selected from the group contained in Table 2. In another embodiment, the polymorphic site contains a genetic variant, and in particular, the genetic variants contained in Table 2 or the complements of the variants in Table 2. In yet another embodiment, the polymorphic site, which may or may not also include a genetic variant, is located at the 3' end of the polynucleotide. In still another embodiment, the polynucleotide further contains a detectable marker. Suitable markers include, but are not limited to, radioactive labels, such as radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.

The present invention also includes kits for the detection of polymorphisms associated with diseases, conditions or disorders, and in particular end-stage renal disease and hypertension. The kits contain, at a minimum, at least one polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO 1 or 2, or the complements of SEQ ID NO: 1 or 2. In one embodiment, the polynucleotide contains at least one polymorphic site, preferably a polymorphic site selected from the group contained in Table 2. Alternatively the 3' end of the polynucleotide is immediately 5' to a polymorphic site, preferably a polymorphic site contained in Table 2. In one embodiment, the polymorphic site contains a genetic variant, preferably a genetic variant selected from the group contained in Table 2. In still another embodiment, the genetic variant is located at the 3' end of the polynucleotide. In yet another embodiment, the polynucleotide of the kit contains a detectable label. Suitable labels include, but are not limited to, radioactive labels, such as radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.

In addition, the kit may also contain additional materials for detection of the polymorphisms. For example, and without limitation, the kits may contain buffer

solutions, enzymes, nucleotide triphosphates, and other reagents and materials necessary for the detection of genetic polymorphisms. Additionally, the kits may contain instructions for conducting analyses of samples for the presence of polymorphisms and for interpreting the results obtained.

In yet another embodiment the present invention provides a method for designing a treatment regime for a patient having a disease, condition or disorder and in particular end stage renal disease and hypertension caused either directly or indirectly by the presence of one or more single nucleotide polymorphisms. In this method genetic material from a patient, for example, DNA, cDNA, RNA or mRNA is screened for the presence of one or more SNPs associated with the disease of interest. Depending on the type and location of the SNP, a treatment regime is designed to counteract the effect of the SNP.

Alternatively, information gained from analyzing genetic material for the presence of polymorphisms can be used to design treatment regimes involving gene therapy. For example, detection of a polymorphism that either affects the expression of a gene or results in the production of a mutant protein can be used to design an artificial gene to aid in the production of normal, wild type protein or help restore normal gene expression. Methods for the construction of polynucleotide sequences encoding proteins and their associated regulatory elements are well know to those of ordinary skill in the art. Once designed, the gene can be placed in the individual by any suitable means known in the art (*Gene Therapy Technologies, Applications and Regulations*, Meager, ed., Wiley, 1999; *Gene Therapy: Principles and Applications*, Blankenstein, ed., Birkhauser Verlag, 1999; Jain, *Textbook of Gene Therapy*, Hogrefe and Huber, 1998).

The present invention is also useful in designing prophylactic treatment regimes for patients determined to have an increased susceptibility to a disease, condition or disorder, and in particular end stage renal disease and hypertension due to the presence of one or more single nucleotide polymorphisms. In this embodiment, genetic material, such as DNA, cDNA, RNA or mRNA, is obtained from a patient and screened for the presence of one or more SNPs associated either directly or indirectly to a disease, condition, disorder or other pathological condition. Based on this information, a treatment regime can be designed to decrease the risk of the patient developing the disease. Such treatment can include, but is not limited to, surgery, the administration of pharmaceutical compounds or nutritional

supplements, and behavioral changes such as improved diet, increased exercise, reduced alcohol intake, smoking cessation, etc.

EXAMPLES

Positions of the single nucleotide polymorhisms (SNPs) are given according to the numbering scheme in the GenBank Accession No. J04431 for the TGF-b1 promoter region and Accession No. U37070 for the TGF-b1 receptor II promoter region. GenBank sequences can be found at http://www.ncbi.nlm.nih.gov/

In the following examples, SNPs are written as "wild type nucleotide" --> "variant nucleotide." Changes in nucleotide sequences are indicated in bold print and with double underlining. The standard nucleotide abbreviations are used in which A=adenine, C=cytosine, G=guanine, T=thymine, M=A or C, R=A or G, W=A or T, S=C or G, Y=C or T, K=G or T, V=A or C or G, H=A or C or T; D=A or G or T; B=C or G or T; N= A or C or G or T.

Example 1

Detection of Novel Polymorphisms by PCR-SSCA

Leukocytes were obtained from human whole blood collected with EDTA. Blood was obtained from a group of 40 patients with ESRD. The group was made up of Caucasians and African Americans of both sexes in unequal ratios.

Genomic DNA was purified from the collected leukocytes using standard protocols well known to those of ordinary skill in the art of molecular biology (Ausubel et al., *Short Protocol in Molecular Biology*, 3rd ed., John Wiley and Sons, 1995; Sambrook et al., *Molecular Cloning*, Cold Spring Harbor Laboratory Press, 1989; and Davis et al., *Basic Methods in Molecular Biology*, Elsevier Science Publishing, 1986). One hundred nanograms of purified genomic DNA was used in each PCR reaction.

Standard PCR reaction conditions were used. Methods for conducting PCR are well known in the art and can be found, for example, in U.S. Patent Nos 4,965,188, 4,800,159, 4,683,202, and 4,683,195; Ausbel et al., eds., *Short Protocols in Molecular Biology*, 3rd ed., Wiley, 1995; and Innis et al., eds., *PCR Protocols*, Academic Press, 1990. Specific primers used are given in the following examples.

PCR reactions were carried out in a total volume of 25 ml containing 100 ng leukocyte genomic DNA, 10 pmol of each primer, 200 nM deoxynucleotide

triphosphates (dNTPs), 1 U Taq polymerase (Perkin-Elmer), 1X PCR buffer (50 mM KCL, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, and 0.01% (w/v) gelatin), and 3% (v/v) DMSO. After an initial 3 minutes denaturation at 94°C, 35 cycles were carried out consisting of 1 minute denaturation at 94°C, 1 minute hybridization at 55 °C, 2 minute extension at 72 °C, followed by a final extension step of 5 minute at 72 °C, and 1 minute cooling at 35 °C. Following amplification, 25ml of PCR product were mixed with 12.5 ml denatuaring solution (95% (v/v) formamide, 20 mM xylene cyanole FF) and were incubated for 5 minutes at 95 °C. Subsequently, samples were chilled in an ice-water bath for 5 minutes and loaded onto a 10% (w/v) polyacrylamide gel (acrylamide:bisacrylamide = 49:1; 72 mm X 102 mm X 0.75 mm, Bio-Rad Mini PROTEAN II), containing 0.5X Tris-buffered EDTA solution. Electrophoresis for 17-17.5 hours was carried out under two conditions: 30 V at 4 °C, and 30 V at room temperature (approx. 23 °C). After electrophoresis, gels were stained with 0.5 mg/ml ethidium bromide for 8 minutes.

Any polymorphic bands detected were cut out of the gel, amplified with the same primers as used for that gel's PCR reaction, blunt-end ligated into a vector such as pBlueScript II Ò (Stratagene), cloned, and sequenced. Methods for ligation, cloning and sequencing are well known in the art and can be found, for example in Ausubel et al., *Short Protocol in Molecular Biology*, 3rd ed., John Wiley and Sons, 1995; Sambrook et al., *Molecular Cloning*, Cold Spring Harbor Laboratory Press, 1989; and Davis et al., *Basic Methods in Molecular Biology*, Elsevier Science Publishing, 1986.

Example 2

G to A Substitution at Position 563 of Human TGF-b1 Promoter

PCR and sequencing were conducted as described in Example 1. The sense primer was 5'-TGCATGGGGACACCATCTACAG-3' (SEQ ID NO: 3) and the antisense primer was 5'-TCTTGACCACTGTGCCATCCTC-3' (SEQ ID NO: 4). The 202 nt PCR product spanned positions 421 to 622 of the human TGF b1 gene (SEQ ID NO: 1). A G563-->A substitution mutation (transition) was found at position 563 with an allele frequency of 33.3% in Caucasian men with ESRD due to HTN but not in a control group of 40 Caucasians and African Americans or in patients with ESRD due to NIDDM.

The polymorphism results in disruption of a potential TCF11 homodimer site (5'-GTCATNNWNNNN-3') (SEQ ID NO: 5) beginning at position 563 on the (+) strand. The indicated G is replaced by an A in this SNP. TCF11 sites occur relatively frequently: 4.63 matches per 1000 base pairs of random genomic DNA sequence in vertebrates.

Other names for TCF11 are KCR-F1 and Nrf1. TCF11 is a transcriptional activator of the basic region-leucine zipper class, abbreviated as "bZIP." It is located on chromosome 17q22, and is homologous to transcription factors involved in Drosophila ("cap 'n collar," or CNC) and C. elegans (skin-1) embryonic development, as well as NF-E2, which regulates the globin locus in erythroid cells. TCF11 has not yet been implicated in control of TGF-b1. Disruption of this transcriptional regulatory site can result in lower transcription of TGF-b1, which runs counter to the currently prevailing notion that TGF-b1 overexpression, rather than underexpression, is associated with chronic renal failure.

The polymorphism also causes disruption of a potential ATF (activating transcription factor) site, which consists of the complement of 5'-GRNNNACGTCASNG-3' (SEQ ID NO: 6), ending at nucleotide 556 on the (-) strand. ATF sites occur relatively rarely: 0.34 times per 1000 base pairs of random genomic sequence in vertebrates. Disruption of this site is also expected to result in decreased transcription of TGF-b1, leading to an expected decrease in the levels of TGF-b1 mRNA and protein in tissues.

Also disrupted is a potential CREB (cAMP-responsive element binding protein) site. Five variations of this site, all centered at this SNP and requiring G563 for maximal activity, exist. In all of them, the indicated G is replaced by an A with this SNP, as follows:

(1) 5'-NSTKACGTCASN-3' (SEQ ID NO: 7), beginning at nucleotide 557; this sequence occurs only 0.09 times per 1000 base pairs of random genomic sequence in vertebrates, so its disruption by this SNP is highly significant.

(2) 5'-NNTTACKGTCASN-3' (SEQ ID NO: 8), also beginning at nucleotide 557; this sequence occurs 0.34 times per 1000 base pairs of random genomic sequence in vertebrates, which is also relatively rare.

(3) 5'-NNTKAC<u>G</u>TCASN-3 (SEQ ID NO: 9), also beginning at nucleotide 557; this sequence occurs 0.34 times per 1000 base pairs of random genomic sequence in vertebrates, which is also relatively rare.

(4) 5'-NNRC<u>G</u>TCANCNN-3' (SEQ ID NO: 10), beginning at nucleotide 559; this sequence occurs less rarely, at 1.12 times per 1000 base pairs of random genomic sequence in vertebrates.

(5) 5'-TKAC<u>G</u>TCA-3', also beginning at nucleotide 559; this sequence occurs 0.40 times per 1000 base pairs of random genomic sequence in vertebrates, which is relatively rare.

There is also disruption of a potential CREBCJUN (cAMP-responsive element binding protein/c-Jun heterodimer) binding site, 5'-TRAC<u>G</u>TCA-3', beginning at nucleotide 559. This sequence occurs relatively rarely in vertebrates, 0.22 times per 1000 base pairs of random genomic sequence.

Additionally, the G563-->A SNP disrupts an AP1 (activator protein 1) site, 5'-WNKNA<u>G</u>TCASY-3' (SEQ ID NO: 11), beginning at nucleotide 558 where the indicated <u>G</u> is replaced with an A. This site occurs relatively frequently: 1.82 times per 1000 bases of random genomic sequence in vertebrates.

The G563-->A SNP appears to be strongly associated with end-stage renal disease due to hypertension in Caucasian men as a major allele. This SNP is predicted to disrupt a number of rare sites for potent transcriptional activators such as TCF11, CRE binding protein, CREB/c-Jun heterodimer, and AP1 (c-Fos/c-Jun heterodimer), as well as more common sites, e.g. for TCF11. The expected result is a decrease in TGF-b1 transcription, with a resultant decrease in tissue abundance of TGF-b1 protein.

Example 3

<u>C to G Substitution at Position 216 of Human TGF-b1 Promoter</u>

PCR and sequencing were conducted as described in Example 1. The sense primer was 5'-AAGACGGTGGGAGCCTAGAAAG-3' (SEQ ID NO: 12) and the antisense primer was 5'-TGGGACCACACCTGGAAATG-3' (SEQ ID NO: 13). The PCR product produced spanned positions 66 to 265 of the human TGF-b1 gene (SEQ ID NO: 1). A C216-->G substitution mutation (transversion) was found at position 216 with an allele frequency of 8.3% in African American women with

ESRD due to NIDDM but not in a control group of 40 Caucasians and African Americans or in patients with ESRD due to HTN..

The polymorphim results in disruption of a putative FSE2 site (nucleotides 216 to 224) in the TGF-b1 promoter, approximately 2kb upstream (5') of the transcription initiation site. The TGF-b1 promoter has two FSE2 sites; the second one is located approximately 600 bases downstream from the first site (at nucleotides 807-816). FSE2 sites are potent negative transcriptional regulatory sites; disruption of a site is thus expected to result in increased transcription of the TGF-b1 gene. Assuming that mRNA stability, translational efficiency, etc. are unchanged, this SNP is expected to result in increased cellular production and secretion of TGF-b1. Since several models of chronic renal failure are associated with increased expression of TGF-b1, this SNP can account for the increased fibrosis of the renal parenchyma observed in chronic renal failure.

The polymorphism also results in disruption of a potential GKLF (gut-enriched Krueppel-like factor) site beginning at nucleotide 211 according to numbering on the (+) strand. The binding site is actually located on the (-) strand, and consists of the complement to the sequence 5'-CCYYTYYYTYNTTY-3' (SEQ ID NO: 14). This SNP replaces the underlined Y (C or T) with a G. GKLF sites occur relatively frequently, 4.76 matches per 1000 base pairs of random genomic sequence in vertebrates.

GKLF is a transcriptional activator, so disruption of its binding site in the TGF-b1 promoter should result in a lower rate of TGF-b1 transcription, and ultimately a lower level of TGF-b1 produced in tissues. However, since GKLF has not been implicated so far in the regulation of TGF-b1 in the kidney, GKLF may represent a heretofore unrecognized participant in TGF-b1 regulation, and hence renal failure.

Example 4

C to T Substitution at Position 552 of TGF-b-RII Promoter

PCR and sequencing were conducted as in Example 1. The sense primer was 5'-AAGTCACTCCAGCTTTGGCAAG-3' (SEQ ID NO: 15) and the antisense primer was 5'-TTGCAGAGGCAGGGTTTGTG-3' (SEQ ID NO: 16). The PCR product produced spanned bases 494-767 of the TGF-b-RII promoter. A C552-->T substitution mutation (transition) was found at position 552 with an allele frequency

of 8.3% in African American women with ESRD due to NIDDM, but not in the reference sequence (SEQ ID NO:2) or patients with ESRD due to HTN.

The polymorphim results in disruption of a potential CETS1P54 [c-Ets-1(p54)] site, which consists of the complement of 5'-NRCWTC<u>C</u>KGN-3' (SEQ ID NO: 17) beginning at position 546. The site is located on the (-) strand. CETS1P54 sites occur relatively frequently: 1.53 matches per 1000 base pairs of random genomic sequence in vertebrates. Since c-Ets-1(p54) is a transcriptional activator, disruption of its binding site is expected to result in a decreased rate of transcription of the TGFb-RII gene. If the rate of translation is tied to the level of messenger RNA, as is the case for most proteins, then less gene product (TGFb-RII) will be the result. TGF-b1 binds to TGFb-RII at the cell surface. TGFb-RII is required for specific recognition of the TGF-b1 ligand. Once the ligand is bound, TGFb-RII then recruits TGFb-RI which is already present (and constitutively active) in the membrane. The heterodimer transduces the TGF-b1 signal to the cell's nucleus through additional components such as the Smad proteins.

Decreased TGFb-RII production is expected to decrease signalling by TGF-b1. How decreased (rather than increased) TGF-b1 signalling leads to ESRD is unclear, and contradicts our current understanding of the pathogenesis of chronic renal failure.

Example 5

<u>Deletion at Position 607 of TGF –b-RII Promoter</u>

PCR and sequencing were conducted as in Example 1. The PCR primers used were the same as in Example 4. A deletion polymorphism in which one of the As at positions 607 and 608 is eliminated was found with an allele frequency of 16.7% in African American women with ESRD due to NIDDM, but not in the reference sequence (SEQ ID NO: 2) or in patients with ESRD due to HTN.

The polymorphism results in disruption of a potential SOX5 binding site, which consists of the sequence 5'-NNAAC<u>A</u>ATNN-3' (SEQ ID NO: 18), beginning at nucleotide 602 on the (+) strand. The A607del polymorphism results in deletion of the indicated <u>A</u>. SOX5 is a member of a large family of transcription factors which bind the so-called "HMG" box, AACAAT. SOX stands for "SRY-related HMG box." SOX5 binding sites occur fairly frequently in vertebrate genomic DNA at 1.10 sites per 1000 bases of random genomic sequence in vertebrates. The SOX

proteins have been implicated in a variety of developmental processes, such as chondrogenesis, but not in the progression of chronic renal failure. Since SOX5 is a transcriptional activator, disruption of its binding site is expected to result in decreased TGFb-RII gene expression.

The polymorphism also causes disruption of a potential TH1E47 (Thing1/E47 heterodimer) binding site, consisting of the sequence complementary to 5'-A<u>A</u>WKCCAGAYNCNNNN-3' (SEQ ID NO: 19) on the (-) strand. The indicated <u>A</u> is deleted by this SNP. TH1E47 binding sites occur relatively frequently, 2.04 sites per 1000 bases of random genomic DNA in vertebrates. Disruption of this site is similarly expected to result in decreased TGFb-RII gene expression.

Example 6

<u>Deletion at Position 619 of TGF-b-RII Promoter</u>

PCR and sequencing were conducted as in Example 1. The PCR primers used were the same as in Example 4. A deletion polymorphism in which one of the Gs at positions 619 and 620 is eliminated was found with an allele frequency of 16.7% in African American women with ESRD due to NIDDM, but not in the reference sequence (SEQ ID NO: 2) or in patients with ESRD sue to HTN.

The G619del SNP disrupts the TH1E47 (Thing1/E47 heterodimer) binding site, which consists of the sequence complementary to 5'-AAWKCCAGAYNCN<u>N</u>NN-3' (SEQ ID NO: 20). This sequence begins at nucleotide 606 on the (+) strand. The TH1E47 binding site is located on the (-) strand. The indicated <u>N</u> is deleted by this SNP, but it is expected to make no difference in binding by TH1E47. TH1E47 binding sites occur relatively frequently, 2.04 sites per 1000 bases of random genomic DNA in vertebrates.

Example 7

<u>Insertion at Position 757 of TGF-b-RII Promoter</u>

PCR and sequencing were conducted as in Example 1. The sense primer was 5'-GGAGTTGGGTTTGGGGGAG-3' (SEQ ID NO: 21) and the antisense primer was 5'-TCTTGCTAGGGCAACCAGATTTG-3' (SEQ ID NO: 22). The PCR product produced spanned bases 697 to 988 of the TGF-b-RII promoter (SEQ ID NO: 2). An insertion polymorphism in which an additional T is inserted at position

757 was discovered with an allele frequency of 8.3% in African American women with ESRD due to NIDDM and 33.3% in African American women with ESRD due to HTN, but not in the reference sequence (SEQ ID NO: 2).

Insertion of a T after nucleotide 757 occurs in a potential NF-1 (nuclear factor-1) site on the (+) strand, consisting of the sequence 5'-NNTTGGCNNNNNNCC<u>N</u>NN-3' (SEQ ID NO: 23), beginning at position 742. T is inserted after the indicated nucleotide <u>N</u>. NF1 binding sites occur rather frequently, 4.11 times per 1000 base pairs of random genomic sequence in vertebrates. However, this insertion is not expected to disrupt binding by NF-1, as there is no stringency for the final three nucleotides.

Example 8

A to G Substitution at Position 792 of TGF-b-RII Promoter

PCR and sequencing were conducted as in Example 1. The primers were the same as in Example 7. An A to G substitution polymorphism (transition) was discovered at position 792 with an allele frequency of 4.2% in African American women with ESRD due to NIDDM and 16.7% in African American women with ESRD due to HTN, but not in the reference sequence (SEQ ID NO: 2).

The polymorphism results in disruption of an S8 binding site, which consists of the sequence 5'-NNNNNYA<u>A</u>TTAN-3' (SEQ ID NO: 24), beginning at position 784 on the (+) strand. This polymorphism replaces the indicated <u>A</u> with a G, which should result in a weaker binding site for S8, a transcriptional activator. S8 binding sites occur rather frequently, 1.73 times per 1000 base pairs of random genomic sequence in vertebrates.

S8 is a homeobox protein, also called Prx2, which operates at the junction of epithelial and mesenchymal tissues. It is a transcription factor active during organogenesis. During development of the kidney, ureteral ectoderm (i.e. epithelial cells) induce surrounding mesoderm (or mesenchyme) to differentiate into proximal and distal tubules. It has been suggested that developmental pathways operating during embryonic life might be reactivated during disease processes.

Since S8 is a positive transcriptional regulator, disruption of its binding site is expected to result in a decreased rate of transcription of the TGFb-RII gene. If the rate of translation is tied to the level of messenger RNA, as is the case for most

proteins, then less gene product (TGFb-RII protein) will be the result, ultimately leading to less TGFb1 signalling in tissues.

The polymorphism also results in disruption of a TCF11 (TCF11/KCR-F1/Nrf1 homodimer) binding site, which consists of the sequence 5'-GTCATNNWNNNNN-3' (SEQ ID NO: 25), beginning at position 789 on the (+) strand. This polymorphism replaces the indicated A with a G, which should result in a weaker binding site for TCF11, a transcriptional activator. TCF11 binding sites occur moderately frequently, at the rate of 4.63 sites per 1000 base pairs of random genomic sequence in vertebrates. Since TCF11 is a transcriptional activator, disruption of its binding site is expected to result in a decreased rate of transcription of the TGFb-RII gene. If the rate of translation is tied to the level of messenger RNA, as is the case for most proteins, then less gene product (TGFb-RII protein) will be the result, ultimately leading to less TGFb1 signalling in tissues.

The A792-->G SNP appears to be a relatively important allele for ESRD due to hypertension in African American women.

Example 9

G to A Substitution at Position 1009 of TGF-b-RII Promoter

PCR and sequencing were conducted as in Example 1. The sense primer was 5'-GGACATATCTGAAAGAGAAAGGGGG-3' (SEQ ID NO: 26) and the antisense primer was 5'-TTGGGAGTCACCTGAATGCTTG-3' (SEQ ID NO: 27). The PCR product produced spanned bases 892 to 1113 of the TGF-b-RII promoter. A G1009-->A substitution polymorphism (transition) was discovered at position 1009. This polymorphism had an allele frequency of 37.5% in members of a control group of 20 Caucasians and African Americans In patients with ESRD due to NIDDM, the polymorphism had an allele frequency of 25% in Caucasian men, 12.5% in Caucasian women, 33.3% in African American men, and 58.3% in African American women. In patients with ESRD due to HTN, the polymorphism had an allele frequency of 16.7% in Caucasian men, 25% in Caucasian women, 40% in African American men, and 50% in African American women.

The G1009-->A transition is quite frequent in the control population, with an allele frequency of 37.5%. It appears to occur at roughly the same frequency in African American men with ESRD due to hypertension (40%) or NIDDM (33%), but at a significantly lower frequency in both male and female Caucasians with

ESRD due to hypertension or NIDDM. It appears to occur at a significantly higher frequency among African American women with ESRD due to either hypertension (50%) or NIDDM (58%).

These data would suggest that the G1009-->A SNP is protective for Caucasians with NIDDM or hypertension, neutral for African American men with NIDDM or hypertension, but harmful for African American women with NIDDM or hypertension. The mechanism for such a complex effect is unknown, but suggests the involvement of additional modifier genes that differ in frequency according to gender and ethnicity.

A comparison of the allele frequency of the G1009-->A SNP was also made between a group of 28 Caucasian males with HTN, but not ESRD, and a control group of 26 Caucasian males with neither HTN nor ESRD. The allele frequency for the G1009-->A SNP was 12.5% in the group with HTN as compared to 34.6% in the control group. These data indicate that the G1009-->A SNP is protective for the development of HTN.

Table 2

Gene	Region	Location	Wild Type	Variant	SEQ ID
TGF-b1	Promoter	216	C	G	1
		563	G	A	1
TGF-b-RII	Promoter	552	C	T	2
		607-608	AA	A	2
		619-620	GG	G	2
		757	T	TT	2
		792	A	G	2
		1009	G	A	2

What is claimed is:

1. A method for diagnosing a genetic susceptibility for end-stage renal disease in a subject comprising:
obtaining a biological sample containing at least one polynucleotide from said subject; and
analyzing said polynucleotide to detect a genetic polymorphism wherein said genetic polymorphism is associated with an altered susceptibility for end-stage renal disease.

2. The method of claim 1, wherein said polynucleotide is DNA, RNA, cDNA or mRNA.

3. The method of claim 1, wherein said polymorphism is located in a gene selected from the group consisting of the genes in Table 1.

4. The method of claim 1, wherein said polymorphism is located in the TGF-b1 gene or the TGF-b-RII gene.

5. The method of claim 1, wherein said genetic polymorphism is located at a site selected from the group consisting of the polymorphic sites in Table 2.

6. The method of claim 1, wherein said genetic polymorphism is a genetic variant selected from the group consisting of the variants in Table 2.

7. The method of claim 1, wherein said genetic polymorphism is a genetic variant selected from the group consisting of the complements of the genetic variants in Table 2.

8. The method of claim 1, wherein said analysis is accomplished by sequencing, mini sequencing, hybridization, restriction fragment analysis, oligonucleotide ligation assay or allele specific PCR.

9. An isolated nucleic acid sequence comprising at least 10 contiguous nucleotides from SEQ ID NO: 1 or 2, wherein said sequence contains at least one variant associated with end stage renal disease so that said sequence differs from SEQ ID NO: 1 or 2.

10. The isolated nucleic acid sequence of claim 9, wherein said variant is located at a site selected from the group consisting of the sites in Table 2.

11. The isolated nucleic acid sequence of claim 9, wherein said variant is selected from the group consisting of the variants in Table 2.

12. The isolated nucleic acid sequence of claim 9, wherein said variant is selected from the group consisting of the complements of the variants in Table 2.

13. The isolated sequence of claim 9, wherein said variant is located at the 3' end of said nucleic acid sequence.

14. The isolated nucleic acid sequence of claim 9, further comprising a detectable label.

15. The isolated nucleic acid sequence of claim 14, wherein said detectable label is selected from the group consisting of radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.

16. A kit for the detection of a polymorphism comprising at least one polynucleotide of at least 10 contiguous nucleotides from SEQ ID NO 1 or 2, wherein said polynucleotide contains at least one variant associated with end stage renal disease; a means for detecting the presence of said polymorphism; and instructions for detecting said polymorphism.

17. The kit of claim 16, wherein said variant is located at a site selected from the group consisting of the polymorphic sites in Table 2.

18. The kit of claim 16, wherein said variant is selected from the group consisting of the variants in Table 2.

19. The kit of claim 16, wherein said variant is selected from the group consisting of the complements of the variants in Table 2.

20. The kit of claim 16, wherein said variant is locatred at the 3' end of said polynucleotide.

21. The kit of claim 16, wherein said polynucleotide further comprises a detectable label.

22. The kit of claim 21, wherein said detectable label is chosen from the group consisting of radionuclides, fluorophores or fluorochromes, peptides enzymes, antigens, antibodies, vitamins or steroids.

23. A kit for the detection of a polymorphism comprising at least one polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO: 1 or 2 or the complement thereof; wherein the 3' end of said polynucleotide is immediately 5' to a genetic variant chosen from the group consisting of the variants in Table 2 and the complements thereof.

24. The kit of claim 23, wherein said polynucleotide further comprises a detectable label.

25. The kit of claim 24, wherein said detectable label is chosen from the group consisting of radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.

26. A method for treatment or prophylaxis in a subject with a polymorphism associated with end stage renal disease comprising:
obtaining a sample of biological material containing at least one polynucleotide from a subject;

analyzing said polynucleotide to detect the presence of at least one polymorphism associated with end stage renal disease; and
treating said subject.

27. The method of claim 26 wherein said polynucleotide is selected from the group consisting of DNA, cDNA, RNA and mRNA.

28. The method of claim 26, wherein said polymorphsim is located in a gene selected from the group consisting of the genes in Table 1.

29. The method of claim 26, wherein said polymorophism is located in the TGF-b1 gene or the TGF-b-RII gene.

30. The method of claim 26, wherein said polymorphism is a variant selected from the group consisting of the variants in Table 2 or the complements thereof.

31. The method of claim 26 wherein said treatment counteracts the effect of the polymorphism.

32. A method for diagnosing a genetic susceptibility for hypertension in a subject comprising:
obtaining a biological sample containing at least one polynucleotide from said subject; and
analyzing said polynucleotide to detect a genetic polymorphism wherein said genetic polymorphism is associated with an altered susceptibility for end-stage renal disease.

33. The method of claim 32, wherein said polynucleotide is DNA, RNA, cDNA or mRNA.

34. The method of claim 32, wherein said polymorphism is located in a gene selected from the group consisting of the genes in Table 1.

35. The method of claim 32, wherein said polymorphism is located in the TGF-b-RII gene.

36. The method of claim 35, wherein said genetic polymorphism is located at position 1009.

37. The method of claim 36, wherein said genetic polymorphism is a G to A substitution.

38. The method of claim 36, wherein said genetic polymorphism is a C to T substitution.

39. The method of claim 32, wherein said analysis is accomplished by sequencing, mini sequencing, hybridization, restriction fragment analysis, oligonucleotide ligation assay or allele specific PCR.

40. A method for treatment or prophylaxis in a subject with a polymorphism associated with hypertension comprising:
obtaining a sample of biological material containing at least one polynucleotide from a subject;
analyzing said polynucleotide to detect the presence of at least one polymorphism associated with end stage renal disease; and
treating said subject.

41. The method of claim 40 wherein said polynucleotide is selected from the group consisting of DNA, cDNA, RNA and mRNA.

42. The method of claim 40, wherein said polymorphsim is located in a gene selected from the group consisting of the genes in Table 1.

43. The method of claim 40, wherein said polymorophism is located in the TGF-b-RII gene.

44. The method of claim 43 wherein said polymorphism is located at position 1009.

45. The method of claim 44, wherein said polymorphism is a G to A substitution

46. The method of claim 44, wherein said polymorphism is a C to T substitution.

47. The method of claim 40 wherein said treatment counteracts the effect of the polymorphism.

FIGURE 1

```
   1 ggatccttag caggggagta acatggattt ggaaagatca ctttggctgc tgtgtgggga
  61 tagataagac ggtgggagcc tagaaaggag gctgggttgg aaactctggg acagaaaccc
 121 agagaggaaa agactgggcc tggggtctcc agtgagtatc agggagtggg gaatcagcag
 181 gagtctggtc cccacccatc cctcctttcc cctctctctc ctttcctgca ggctggcccc
 241 ggctccattt ccaggtgtgg tcccaggaca gctttggccg ctgccagctt gcaggctatg
 301 gattttgcca tgtgcccagt agcccgggca cccaccagct ggcctgcccc acgtggcggc
 361 ccctgggcag ttggcgagaa cagttggcac gggctttcgt gggtggtggg ccgcagctgc
 421 tgcatgggga caccatctac agtggggccg accgctatcg cctgcacaca gctgctggtg
 481 gcaccgtgca cctggagatc ggcctgctgc tccgcaactt cgaccgctac ggcgtggagt
 541 gctgagggac tctgcctcca acgtcaccac catccacacc ccggacaccc agtgatgggg
 601 gaggatggca cagtggtcaa gagcacagac tctagagact gtcagagctg accccagcta
 661 aggcatggca ccgcttctgt cctttctagg acctcggggt ccctctgggc ccagtttccc
 721 tatctgtaaa ttggggacag taaatgtatg gggtcgcagg gtgttgagtg acaggaggct
 781 gcttagccac atgggaggtg ctcagtaaag gagagcaatt cttacaggtg tctgcctcct
 841 gacccttcca tccctcaggt gtcctgttgc cccctcctcc cactgacacc ctccggaggc
 901 ccccatgttg acagaccctc cttctcctac cttgtttccc agcctgactc tccttccgtt
 961 ctgggtcccc ctcctctggt cggctccect gtgtctcatc ccccggatta agccttctcc
1021 gcctggtcct ctttctctgg tgacccacac cgcccgcaaa gccacagcgc atctggatca
1081 cccgctttgg tggcgcttgg ccgccaggag gcagcaccct gtttgcgggg cggagccggg
1141 gagcccgccc cctttccccc agggctgaag ggaccccccc cggagcccgc ccacgcgaga
1201 tgaggacggt ggcccagccc ccccatgccc tccccctggg ggccgccccc gctcccgccc
1261 cgtgcgcttc ctgggtgggg ccgggggcgg cttcaaaacc ccctgccgac ccagccggtc
1321 cccgccgccg ccgcccttcg cgccctgggc catctccctc ccacctccct ccgcggagca
1381 gccagacagc gagggccccg gccgggggca gggggacgc cccgtccggg gcaccccccc
1441 ggctctgagc cgcccgcggg gccggcctcg gcccggagcg gaggaaggag tcgccgagga
1501 gcagcctgag gccccagagt ctgagacgag ccgccgccgc ccccgccact gcggggagga
1561 gggggaggag gagcgggagg agggacgagc tggtcgggag aagaggaaaa aaacttttga
1621 gacttttccg ttgccgctgg gagccggagg cgcggggacc tcttggcgcg acgctgcccc
1681 gcgaggaggc aggacttggg gaccccagac cgcctccctt tgccgccggg gacgcttgct
1741 ccctccctgc cccctacacg gcgtccctca ggcgcccccа ttccggacca gccctcggga
1801 gtcgccgacc cggcctcccg caaagacttt tccccagacc tcgggcgcac cccctgcacg
1861 ccgccttcat ccccggcctg tctcctgagc ccccgcgcat cctagaccct ttctcctcca
1921 ggagacggat ctctctccga cctgccacag atcccctatt caagaccacc caccttctgg
1981 taccagatcg cgcccatcta ggttatttcc gtgggatact gagacacccc cggtccaagc
2041 ctcccctcca ccactgcgcc cttctccctg aggagcctca gctttccctc gaggccctcc
2101 taccttttgc cgggagaccc ccagcccctg caggggcggg gcctccccac cacaccagcc
2161 ctgttcgcgc tctcggcagt gccgggggggc gccgcctccc ccatg
```

FIGURE 2

```
   1 cccatcaaag aagttatgat tcaatccacg aagaccagga gttggcgaaa tgaagaaaaa
  61 aaggtcagag gaaggaagtc ctctctgggg aaggctctaa gcataaaggg caggaggatt
 121 acagaggcat atctcgaaat ttggagaagg ctttcagtaa gcaaggagaa gccaaatgaa
 181 agtttacgga gagttggagg cttgaagaca ccgttcaagg atctggtttt tatcttctct
 241 ttattctcaa gagcttagtg ggaagccatt aaatgatttt aatcaaggag gggttggtta
 301 taaactagtt ttgttaattt tgaaaaatct gaattcactc tcgtttgaga aactgagtga
 361 aagagcccag aacggccgtg ctgagggtga ctcctgggaa gactccttaa ccacaagcca
 421 tggcagtggc atgggctggt ggcagaagag ggaataggga gaagatttgg aactcaatct
 481 tcctccattg acaaagtcac tccagctttg gcaaggcaat taattggtgg gaaagaagat
 541 gcctagccct cctgatttca ctgcactttc tgcatcttca acatgagtac tgggaagtgg
 601 caaaacaatc cagaggcagg cttgggtgct aggtggagca tgagttaaaa ttccaggatg
 661 aagcaaatga acacttagaa tgacaggaaa gatttgggag ttgggtttgg gggagggcta
 721 tttaccttta ttccctggag accctggcac aaaccctgcc tctgcaatct tcctctcagg
 781 taaaggaatt cattaaatga attgctagaa gatctactga ccagagggct gtacagaatc
 841 atatctttga gagtgggaag taggttgatc acatagttta ttatccaatc aggacatatc
 901 tgaaagagaa agggggttct attaatattt aaactacaaa acatgtacac caggaatgtc
 961 ttgggcaaat ctggttgccc tagcaagaaa ggaaatttga aagtttatgc tgttctgctc
1021 ccatgttacc ccgtttgcac atgagagggt aagtattctc tttcttcacc tgcattaagg
1081 gaataaaagc acaagcattc aggtgactcc caacccactt ttaattttac agtttctgct
1141 atactctata cattctgaaa attacatttc ccaccactat acttcgtgat aggtgatcat
1201 ttacaattac tcactgactc agtcccggga agaggcggtg caaaatggac gctctatcca
1261 ggtgctcatt agaaatgcag aatctctgcc tgcctcctag acctactgaa ttagaatctg
1321 catttttaaa taagatttcc aggtgatcaa tatgtacatt aaaacttgag aaaaacctct
1381 agacttcgac ctaaagaaaa acattttaca acttgacagt gtatgcacat acatacatgc
1441 atatagacac aactgaagca caaatttaat gaagtagaat ttaccgttac tattttattt
1501 ggaaagaaat gtgctcgcga ctcaatagat tggagtattc actcctggat ctcaacttgc
1561 aatttgaaaa cgcatctcta aagcacctag gagcaatctg aagaaagctg aggggaggcg
1621 gcagatgttc tgatctacta gggaaaacgt ggacgttttc tgttgttact ttgtgaactg
1681 tgtgcactta gtcattcttg agtaaatact tggagcgagg aactcctgag tggtgtggga
1741 gggcggtgag gggcagctga aagtcggcca aagctctcgg aggggctggt ctaggaaaca
1801 tgattggcag ctacgagaga gctaggggct ggacgtcgag gagagggaga aggctctcgg
1861 gcggagagag gtcctgccca gct
```

SEQUENCE LISTING

<110> Moskowitz, David W

<120> DIAGNOSTIC POLYMORPHISMS

<130> DZG 2176

<140>
<141>

<150> 60/120,787
<151> 1999-02-19

<160> 27

<170> PatentIn Ver. 2.1

<210> 1
<211> 2205
<212> DNA
<213> Homo sapiens

<400> 1
ggatccttag caggggagta acatggattt ggaaagatca ctttggctgc tgtgtgggga 60
tagataagac ggtgggagcc tagaaaggag gctgggttgg aaactctggg acagaaaccc 120
agagaggaaa agactgggcc tggggtctcc agtgagtatc agggagtggg gaatcagcag 180
gagtctggtc cccacccatc cctcctttcc cctctctctc ctttcctgca ggctggcccc 240
ggctccattt ccaggtgtgg tcccaggaca gctttggccg ctgccagctt gcaggctatg 300
gattttgcca tgtgcccagt agcccgggca cccaccagct ggcctgcccc acgtggcggc 360
ccctgggcag ttggcgagaa cagttggcac gggctttcgt gggtggtggg ccgcagctgc 420
tgcatgggga caccatctac agtggggccg accgctatcg cctgcacaca gctgctggtg 480
gcaccgtgca cctggagatc ggcctgctgc tccgcaactt cgaccgctac ggcgtggagt 540
gctgagggac tctgcctcca acgtcaccac catccacacc ccggacaccc agtgatgggg 600

gaggatggca cagtggtcaa gagcacagac tctagagact gtcagagctg accccagcta 660
aggcatggca ccgcttctgt cctttctagg acctcggggt ccctctgggc ccagtttccc 720
tatctgtaaa ttggggacag taaatgtatg gggtcgcagg gtgttgagtg acaggaggct 780
gcttagccac atgggaggtg ctcagtaaag gagagcaatt cttacaggtg tctgcctcct 840
gaccccttcca tccctcaggt gtcctgttgc cccctcctcc cactgacacc ctccggaggc 900
ccccatgttg acagacccctc cttctcctac cttgtttccc agcctgactc tccttccgtt 960
ctgggtcccc ctcctctggt cggctcccct gtgtctcatc ccccggatta agccttctcc 1020
gcctggtcct ctttctctgg tgacccacac cgcccgcaaa gccacagcgc atctggatca 1080
cccgctttgg tggcgcttgg ccgccaggag gcagcaccct gtttgcgggg cggagccggg 1140
gagcccgccc cctttccccc agggctgaag ggaccccccct cggagcccgc ccacgcgaga 1200
tgaggacggt ggcccagccc ccccatgccc tcccccctggg ggccgccccc gctcccgccc 1260
cgtgcgcttc ctgggtgggg ccgggggcgg cttcaaaacc ccctgccgac ccagccggtc 1320
cccgccgccg ccgcccttcg cgccctgggc catctccctc ccacctccct ccgcggagca 1380
gccagacagc gagggccccg gccgggggca ggggggacgc cccgtccggg gcaccccccc 1440
ggctctgagc cgcccgcggg gccggcctcg gcccggagcg gaggaaggag tcgccgagga 1500
gcagcctgag gccccagagt ctgagacgag ccgccgccgc ccccgccact gcggggagga 1560
gggggaggag gagcgggagg agggacgagc tggtcgggag aagaggaaaa aaacttttga 1620
gacttttccg ttgccgctgg gagccggagg cgcgggggacc tcttggcgcg acgctgcccc 1680
gcgaggaggc aggacttggg gaccccagac cgcctccctt tgccgccggg gacgcttgct 1740
ccctccctgc cccctacacg gcgtccctca ggcgcccccа ttccggacca gccctcggga 1800
gtcgccgacc cggcctcccg caaagacttt tccccagacc tcgggcgcac cccctgcacg 1860
ccgccttcat ccccggcctg tctcctgagc ccccgcgcat cctagaccct ttctcctcca 1920
ggagacggat ctctctccga cctgccacag atcccctatt caagaccacc caccttctgg 1980
taccagatcg cgcccatcta ggttatttcc gtgggatact gagacacccc cggtccaagc 2040
ctcccctcca ccactgcgcc cttctccctg aggagcctca gctttccctc gaggccctcc 2100
taccttttgc cgggagaccc ccagcccctg caggggcggg gcctccccac cacaccagcc 2160
ctgttcgcgc tctcggcagt gccgggggggc gccgcctccc ccatg 2205

<210> 2
<211> 1883
<212> DNA
<213> Homo sapiens

<400> 2

cccatcaaag aagttatgat tcaatccacg aagaccagga gttggcgaaa tgaagaaaaa 60
aaggtcagag gaaggaagtc ctctctgggg aaggctctaa gcataaaggg caggaggatt 120
acagaggcat atctcgaaat ttggagaagg ctttcagtaa gcaaggagaa gccaaatgaa 180
agtttacgga gagttggagg cttgaagaca ccgttcaagg atctggtttt tatcttctct 240
ttattctcaa gagcttagtg ggaagccatt aaatgatttt aatcaaggag gggttggtta 300
taaactagtt ttgttaattt tgaaaaatct gaattcactc tcgtttgaga aactgagtga 360
aagagcccag aacggccgtg ctgagggtga ctcctgggaa gactccttaa ccacaagcca 420
tggcagtggc atgggctggt ggcagaagag ggaataggga gaagatttgg aactcaatct 480
tcctccattg acaaagtcac tccagctttg gcaaggcaat taattggtgg gaaagaagat 540
gcctagccct cctgatttca ctgcactttc tgcatcttca acatgagtac tgggaagtgg 600
caaaacaatc cagaggcagg cttgggtgct aggtggagca tgagttaaaa ttccaggatg 660
aagcaaatga acacttagaa tgacaggaaa gatttgggag ttgggtttgg gggagggcta 720
tttaccttta ttccctggag accctggcac aaaccctgcc tctgcaatct tcctctcagg 780
taaaggaatt cattaaatga attgctagaa gatctactga ccagagggct gtacagaatc 840
atatctttga gagtgggaag taggttgatc acatagttta ttatccaatc aggacatatc 900
tgaaagagaa agggggttct attaatattt aaactacaaa acatgtacac caggaatgtc 960
ttgggcaaat ctggttgccc tagcaagaaa ggaaatttga aagtttatgc tgttctgctc 1020
ccatgttacc ccgtttgcac atgagagggt aagtattctc tttcttcacc tgcattaagg 1080
gaataaaagc acaagcattc aggtgactcc caacccactt ttaattttac agtttctgct 1140
atactctata cattctgaaa attacatttc ccaccactat acttcgtgat aggtgatcat 1200
ttacaattac tcactgactc agtcccggga agaggcggtg caaaatggac gctctatcca 1260
ggtgctcatt agaaatgcag aatctctgcc tgcctcctag acctactgaa ttagaatctg 1320
catttttaaa taagatttcc aggtgatcaa tatgtacatt aaaacttgag aaaaacctct 1380
agacttcgac ctaaagaaaa acattttaca acttgacagt gtatgcacat acatacatgc 1440
atatagacac aactgaagca caaatttaat gaagtagaat ttaccgttac tattttattt 1500
ggaaagaaat gtgctcgcga ctcaatagat tggagtattc actcctggat ctcaacttgc 1560
aatttgaaaa cgcatctcta aagcacctag gagcaatctg aagaaagctg aggggaggcg 1620
gcagatgttc tgatctacta gggaaaacgt ggacgttttc tgttgttact ttgtgaactg 1680
tgtgcactta gtcattcttg agtaaatact tggagcgagg aactcctgag tggtgtggga 1740
gggcggtgag gggcagctga aagtcggcca aagctctcgg aggggctggt ctaggaaaca 1800
tgattggcag ctacgagaga gctaggggct ggacgtcgag gagagggaga aggctctcgg 1860
gcggagagag gtcctgccca gct 1883

<210> 3
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 3
tgcatgggga caccatctac ag 22

<210> 4
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 4
tcttgaccac tgtgccatcc tc 22

<210> 5
<211> 12
<212> DNA
<213> Unknown Organism

<220>
<221> mutation
<222> (1)
<223> G is replaced by A

<220>
<221> unsure
<222> (6)..(7)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (9)..(12)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential TCFII homodimer

<400> 5
gtcatnnwnn nn 12

<210> 6
<211> 14
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (3)..(5)
<223> N= A or G or C or T

<220>
<221> mutation
<222> (8)
<223> G is replaced by A

<220>

<221> unsure
<222> (13)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential activating transcription factor (ATF) site

<400> 6
grnnnacgtc asng 14

<210> 7
<211> 12
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)
<223> N= A or G or C or T

<220>
<221> mutation
<222> (7)
<223> G is replaced by A

<220>
<221> unsure
<222> (12)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential cAMP-

responsive element binding protein (CREB) site

<400> 7
nstkacgtca sn 12

<210> 8
<211> 13
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)..(2)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (8)
<223> G is replaced by A

<220>
<221> unsure
<222> (13)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential cAMP-responsive element binding protein (CREB) site

<400> 8
nnttackgtc asn 13

<210> 9
<211> 12
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)..(2)
<223> N= A or G or C or T

<220>
<221> mutation
<222> (7)
<223> G is replaced with A

<220>
<221> unsure
<222> (12)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential cAMP-responsive element binding protein (CREB) site

<400> 9
nntkacgtca sn 12

<210> 10
<211> 12
<212> DNA
<213> Unknown Organism

<220>

<221> unsure
<222> (1)..(2)
<223> N= A or G or C or T

<220>
<221> mutation
<222> (5)
<223> G is replaced with A

<220>
<223> Description of Unknown Organism: potential cAMP-responsive element binding protein (CREB) site

<400> 10
nnrcgtcanc nn 12

<210> 11
<211> 11
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (2)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (4)
<223> N= A or G or C or T

<220>
<221> mutation

<222> (6)
<223> G is replaced with A

<220>
<223> Description of Unknown Organism: activator protein 1 (AP1) site

<400> 11
wnknagtcas y 11

<210> 12
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 12
aagacggtgg gagcctagaa ag 22

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 13
tgggaccaca cctggaaatg 20

<210> 14
<211> 14
<212> DNA
<213> Unknown Organism

<220>
<221> mutation
<222> (6)
<223> Y (C or T) is replaced with G

<220>
<223> Description of Unknown Organism: potential gut enriched Krueppel- like factor (GKLF) site

<400> 14
ccyytyyyty ntty 14

<210> 15
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 15
aagtcactcc agctttggca ag 22

<210> 16
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 16
ttgcagaggc agggtttgtg 20

<210> 17
<211> 10
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)
<223> N= A or G or C or T

<220>
<221> mutation
<222> (7)
<223> C is replaced by T

<220>
<221> unsure
<222> (10)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential
CETS1P54 [c-Ets-1(p54)] site

<400> 17
nrcwtcckgn 10

<210> 18
<211> 10
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)..(2)
<223> N= A or G or C or T

<220>
<221> misc_feature
<222> (6)
<223> A is deleted

<220>
<221> unsure
<222> (9)..(10)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential SOX5 binding site

<400> 18
nnaacaatnn 10

<210> 19
<211> 16
<212> DNA
<213> Unknown Organism

<220>

<221> misc_feature
<222> (2)
<223> A is deleted

<220>
<221> unsure
<222> (11)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (13)..(16)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: potential Thing 1/E47 heterodimer (TH1E47) binding site

<400> 19
aawkccagay ncnnnn 16

<210> 20
<211> 16
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (11)
<223> N= A or G or C or T

<220>
<221> unsure

<222> (13)..(16)
<223> N= A or G or C or T

<220>
<221> misc_feature
<222> (14)
<223> N is deleted

<220>
<223> Description of Unknown Organism: Thing 1/E47 heterodimer (TH1E47) binding site

<400> 20
aawkccagay ncnnnn 16

<210> 21
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 21
ggagttgggt ttgggggag 19

<210> 22
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 22
tcttgctagg gcaaccagat ttg 23

<210> 23
<211> 18
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)..(2)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (8)..(13)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (16)..(18)
<223> N= A or G or C or T

<220>
<221> misc_feature
<222> (16)
<223> T is inserted after N

<220>
<223> Description of Unknown Organism: potential nuclear factor-1 (NF-1) site

<400> 23
nnttggcnnn nnnccnnn 18

<210> 24
<211> 12
<212> DNA
<213> Unknown Organism

<220>
<221> unsure
<222> (1)..(5)
<223> N= A or G or C or T

<220>
<221> mutation
<222> (8)
<223> A is replaced with G

<220>
<221> unsure
<222> (12)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: S8 binding site

<400> 24
nnnnnyaatt an 12

<210> 25
<211> 13
<212> DNA

<213> Unknown Organism

<220>
<221> mutation
<222> (4)
<223> A is replaced with G

<220>
<221> unsure
<222> (6)..(7)
<223> N= A or G or C or T

<220>
<221> unsure
<222> (9)..(13)
<223> N= A or G or C or T

<220>
<223> Description of Unknown Organism: TCF11/KCR-F1/Nrf1 homodimer (TCF11) binding site

<400> 25
gtcatnnwnn nnn 13

<210> 26
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 26

ggacatatct gaaagagaaa ggggg 25

<210> 27
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 27
ttgggagtca cctgaatgct tg 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04251

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12P 19/34; C12Q 1/68
US CL :435/91.1, 91.2, 6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91.1, 91.2, 6, 91.51; 436/94: 536/23.1, 24.3, 24.33, 25.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI et al. TGF-beta1 DNA polymorphisms, protein levels, and blood pressure. Hypertension, January 1999, Vol. 33, No. 1, Pages 271-275, especially 271-273.	1-4, 8, 32-34 and 39
Y	FREEDMAN et al. Genetic linkage analysis of growth factor loci and end-stage renal disease in african americans. Kidney International, 1997, Vol. 51, Pages 819-825, especially 819-822.	1-8 and 32-39
Y	PEI et al., Association of angiotensinogen gene T235 variant with progression of immunoglobin a nephropathy in caucasian patients. J. Clin. Invest. August 1997, Vol. 100, No. 4, Pages 814-820, especially pages 814 and 815.	1-8 and 32-39

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 APRIL 2000	07 JUL 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
FRANK LU
Telephone No. (703) 308-1235

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04251

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

STN and WEST
Search term: end-stage renal disease, hypertension, chronic renal failure, polymorphism, transforming growth factor and apoptosis

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-8 and 32-39, drawn to a method for diagnosing a genetic susceptibility for end-stage renal disease (claims 1-8) and a method for diagnosing a genetic susceptibility for hypertension (claims 32-39.

Group II, claims 9-25, drawn to an isolated nucleic acid sequence and a kit for the detection of a polymorphism,

Group III, claims 26-31 and 40-47, drawn to a method for treatment or prophylaxix in a subject with a polymorphism associated with end stage renal disease (claims 26-31) and a method for treatment or prophylaxix in a subject with a polymorphism associated with hypertension.

The inventions listed as Groups I to III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Groups I and II are distinct from each other because these inventions are directed to different products and related methods of use.

Groups I and III are distinct from each other because these inventions are directed to different methods comprised of different method steps and result in different end products.

Groups II and III are distinct from each other because these inventions are directed to different products and related methods of use.